Technische Universität Braunschweig

Institute of Scientific Computing

Oliver Pajonk

Stochastic Spectral Methods for Linear Bayesian Inference

Dissertation

2012

Stochastic Spectral Methods for Linear Bayesian Inference

Von der
Carl-Friedrich-Gauß-Fakultät
der Technischen Universität Carolo-Wilhelmina zu Braunschweig
zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
genehmigte Dissertation

von

Oliver Pajonk
geboren am 8. Februar 1981
in Papenburg

Eingereicht am:	16. Mai 2012
Mündliche Prüfung am:	20. Juli 2012
1. Referent:	Prof. Hermann G. Matthies, Ph.D.
2. Referent:	Prof. Youssef M. Marzouk, Ph.D.

2012

Pajonk, Oliver: Stochastic Spectral Methods for Linear Bayesian Inference

Version: 1.3.2 (Build Version 1634)
This document was created on 2nd August 2012 using LaTeX 2_ε.

Abstract

Simulation-based control of dynamic systems is of key importance for many areas of science and industry. To ensure the predictive capabilities, simulation models used for predicting control responses have to be calibrated to available observations. Bayesian approaches to make inference from data on unobservable quantities are used because of their consistent, inherent treatment of diverse sources of uncertainties.

Spectral approaches to uncertainty quantification have become popular over the last years. However, their combination with Bayesian inference usually employs expensive probabilistic sampling methods. In this work, a family of linear Bayesian approaches is presented which directly results in a representation of the posterior. A specific implementation is discussed which overcomes some of the difficulties that remained unsolved in related approaches. All implementation details are given, and the applicability is demonstrated on some linear and non-linear numerical examples.

Keywords: Numerical mathematics, optimization, scientific computing, system identification, regularization, Bayesian estimation, polynomial chaos expansion, Kalman filter, inverse problem

AMS classification: 60H40, 65M32, 62L12

Zusammenfassung

Die simulationsbasierte Steuerung von dynamischen Systemen stellt eine Schlüsseltechnologie für weite Bereiche von Forschung und Industrie dar. Um die Vorhersagefähigkeiten von Simulationsmodellen sicherzustellen müssen diese auf die verfügbaren Daten kalibriert werden. Bayes'sche Ansätze für die Erzeugung von Rückschlüssen aus Daten auf unbeobachtbare Modellgrößen sind aufgrund ihrer inhärenten Möglichkeiten, Unsicherheiten in den Rückschlussprozess einzubetten, beliebt.

Spektrale Methoden für die Quantifizierung von Unsicherheiten sind über die letzten Jahre populär geworden. Allerdings bedingt ihre Kombination mit Bayes'schen Rückschlussmethoden typischerweise den Einsatz von aufwändigen probabilistischen Abtastverfahren. In dieser Arbeit wird eine Familie von linearen Bayes'schen Vorgehensweisen präsentiert, welche direkt die spektrale à posteriori Repräsentation der unsicheren Zielgröße erzeugen. Eine spezifische Implementierung wird vorgestellt, welche einige der Schwierigkeiten der bisher existierenden Ansätze umgeht. Alle Implementierungsdetails hierzu werden beschrieben, und die Anwendbarkeit anhand von verschiedenen linearen und nicht-linearen numerischen Beispielen belegt.

Contents

List of Figures

List of Algorithms

A beautiful answer to this was given by J. von Neumann in a talk on computers given in Princeton in 1948, which the writer was privileged to attend. In reply to the canonical question from the audience ("But of course, a mere machine can't really think, can it?"), he said: You insist that there is something a machine cannot do. If you will tell me precisely what it is that a machine cannot do, then I can always make a machine which will do just that!

Probability Theory — The Logic of Science, E. T. Jaynes [186, p. 7]

Chapter 1.

Introduction

In many areas of industry and science there is a growing need for computer-based workflows for control under uncertainty with application to dynamical physical systems. A major driving force are large-scale geophysical applications. Such applications are typically highly complex in both their scale (fine scale effects vs. large spatial extent) and behaviour (complex governing equations, for example turbulence models; heterogeneity, for example soil with channels, layers, and fractures).

The key goal is to support decisions which modify the behaviour of the real system in such a way that a quantity of interest attains an optimal value. Examples of such identification and control problems are:

Containment of Groundwater Contamination A contamination source is to be contained in a local area by placing pump stations. For this it is essential to know how the contaminant will spread — which depends crucially on the groundwater flow and thus on subsurface properties like permeability and porosity.

History Matching and Production Optimisation for Oil Reservoirs Oil reservoirs are of a highly complex behaviour and structure. Many of the involved parameters are highly uncertain. Despite these complexities a major interest of the oil industry is to develop methods for maximising the revenue while minimising the risk.

Subsurface CO_2 Sequestration To be able to safely and reliably store CO_2 in subsurface structures one has to have sufficient knowledge about these, especially about sealing layers and faults.

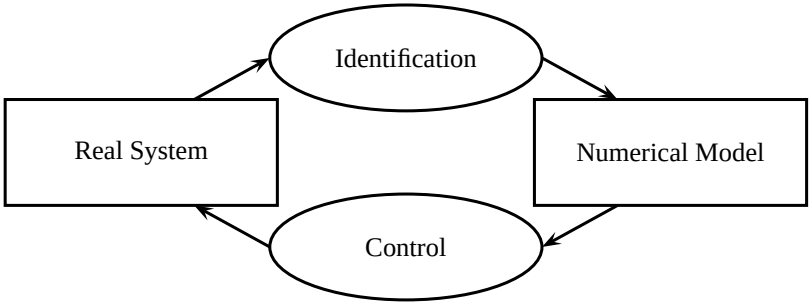

Figure 1.1.: The typical identification and control loop.

Tsunami Modelling Simulating sea-floor motions which could potentially cause tsunamis is the largest unknown in tsunami modelling, and it critically depends on the identification of subsea, subsurface structures from seismic waves and — sometimes — tsunami data ([44, p. 8]). The resulting models are then used for the prediction of tsunami events and by this they are the basis for evacuation control.

Many more examples with similar structure exist in many other areas of application — sometimes with more emphasis on the identification step, sometimes on the control step. Due to the high system complexity, computational methods have to be employed for determining such optimal controls. Naturally, computational methods mainly rely on computer implementations of mathematical models for these large-scale systems, which therefore play a major role in any workflow.

However, it is usual that in mathematical models — because of their naturally limited representation of reality — some important quantities are known only incompletely. Such quantities can be *identified* — up to a certain degree — from observations obtained by performing measurements on the real system. The resulting 'identified' model is hoped to have improved representativeness when compared to the 'un-identified' model since it honours these data. Thus it is clear that a major part of any workflow is necessarily devoted to efficient computational methods for system identification — and this task can be thought of as making inference from incomplete data.

Ultimately, the two steps of identification and control are to be performed in a recursive way, resulting in a closed loop as shown in Fig. 1.1. While here the focus is on the identification part, the control step must always be kept in mind.

The goals of this work are therefore to

- propose a theoretical framework that fits system identification and control tasks for dynamic geophysical and similar systems,
- present an overview of relevant identification methods which fit into this framework and point to related areas of research,
- develop original identification methods or improvements to existing methods where application demands them.

The most important aspect of the framework and all participating methods should be the efficient computational handling of *uncertainty*. Uncertainty at all stages exists due to *incomplete knowledge* of model quantities and various kinds of *errors* in the identification and decision support process, *e.g.* modelling errors, measurement errors and numerical errors. These uncertainties should not be ignored in any identification process, since doing so would likely result in a false sense of security and, subsequently, non-optimal or even disastrous decisions.

1.1. Summary of the State of the Art

The two most important frameworks for the solution of identification problems are regularised optimisation [115] and Bayesian inference [340, 186]. Somewhat related to the Bayesian approach are fuzzy sets and probabilities [74, 75].

Especially the Bayesian view is attracting considerable interest [60]. This is mainly due to its integrated and consistent way to combine various sources of incomplete, uncertain information. Many areas of science and industry begin to realise (or already have realised) that the incorporation of the treatment of uncertainty into their numerical modelling workflows can provide important benefits and is therefore inevitable.

Unfortunately, the treatment of uncertainties is not for free. Methods which push beyond the capabilities of Monte Carlo approaches are necessary — especially when it comes to Bayesian inference. This is a major issue for dynamical systems: a plethora of methods, ranging from Kalman filter variants to highly sophisticated particle filters, have been developed. Combinations of spectral forward modelling and Monte-Carlo-based inference exist. Purely non-probabilistic approaches include the explicit time integration of conditional probabilities and expectations. Only recently first fully spectral approaches to Bayesian inference have appeared and are an important research topic. A review of existing approaches to system identification is given in [1]. Going beyond the identification step, approaches to stochastic optimal control are successfully developed, where prime applications are on autonomous robots.

1.2. Outline

Despite these success stories a significant amount of work remains to be done, and this thesis aims to contribute its small share. Chapter 2 describes a Bayesian framework for inference which — in contrast to most existing ones — uses *random variables* as the basic building block. As an outlook, it is extended towards control applications since this is certainly a major application area for the results of the Bayesian inference. Detailed reviews of existing inference methods can be found in chapter 3. All methods are put into the context of the Bayesian framework, thereby identifying areas where additional research may be worthwhile. The main result of this thesis is given in chapter 4. It introduces a new formulation of a fully spectral method for linear Bayesian inference. By construction, the method is suitable for sequential inference applications. Its implementation is discussed by the example of the *polynomial chaos expansion* (PCE), a spectral representation for random variables. It is shown that this implementation overcomes two major difficulties of existing approaches: (1) compared to existing sequential spectral methods it does not have the problem of a growing spectral basis: no new random variables have to be introduced with every update; and (2) compared to existing sequential sampling approaches it does not involve sampling errors at all — only the truncation error of the spectral representation, which is easier to quantify and control. The method is examined, in comparison to related approaches, on some numerical examples. There it is found to be more

reliable than its sampling-based counterparts: it naturally maintains the span of the stochastic subspace and does not involve sampling at any stage. However, it is also demonstrated that the assumptions involved in the derivation of all these linear Bayesian methods must be considered with care and may have a significant impact on the results, depending on the individual identification problem. Chapter 5 shortly discusses a multi-scale covariance localisation approach based on wavelet analysis. It is intended for probabilistic linear Bayesian methods. Chapter 6 briefly describes a conceptual combination of a population-based evolution strategy with an ensemble-based linear Bayesian method. However, to not obfuscate the main results of this thesis, these two lines of research are discussed only briefly and the reader is referred to the given publications. Finally, chapter 7 concludes the thesis.

Appendix A collects some important methodologies and results on the stochastic representation of uncertainty. Appendix B collects some properties and computational aspects of the PCE which are used throughout the text. Appendix C summarises some metrics which are used as comparison tools in the numerical experiments. Due to the many different ways to implement ensemble Kalman filters, appendix D discusses the variant used for comparison.

1.3. Acknowledgements

First of all I would like to thank Christine for her endurance and support during this thesis, and also the uninterrupted uphold and patience of my parents has been invaluable. It has not always been an easy time with me. My thanks also go to my colleagues, collaborators, and my advisor Hermann G. Matthies for long times of fruitful discussion which ultimately lead to this work. It is impossible to name you all — you know who you are. Last, but most certainly not the least, my thanks go to SPT Group GmbH and Dr. Ralf Schulze-Riegert for their constant support during my work. It would hardly have been possible without you.

Chapter 2.

A Mathematical Framework for Bayesian Inference

In this chapter a generic Bayesian framework for identification under uncertainty is presented. It fixes notation and terms which are then used throughout the work. The framework is related to the one developed by Tarantola [340, chapter 1]. The main difference is the way that is used to represent limited information: here a parametric approach with *random variables* (RVs) as the basic building blocks is used, whereas Tarantola [340] takes a more classical approach and uses probability measures. However, we believe (and try to demonstrate in this chapter) that RVs are more suitable for the task at hand. A second difference is that for us the focus is on *dynamical* systems, which is a basic requirement for the envisioned applications (*cf.* chapter 1). The notation is derived from suggestions in [258, 178] and aims at being as simple as possible.

2.1. System Meta-Model

The system of interest $\mathcal{S}$, which is part of reality, is modelled by a *joint model-data space*

$$\mathcal{M} := \mathcal{X} \times \mathcal{Y}. \tag{2.1}$$

Here $\mathcal{X}$ denotes the *model space* and $\mathcal{Y}$ denotes the *data space*[1]. The space Ω shall capture everything which is missing in $\mathcal{M}$ to actually represent

[1] Details on the subspaces are given in section 2.1.3 on page 11.

$\mathcal{S}$, the reality. This we abstractly write as

$$\mathcal{S} \cong \mathcal{M} \times \Omega. \tag{2.2}$$

The space Ω is discussed later; first focus on $\mathcal{X}$ and $\mathcal{Y}$. These spaces are assumed to be Banach (and sometimes Hilbert) vector spaces[2]. Between the two spaces $\mathcal{X}$ and $\mathcal{Y}$, there exists a mapping

$$\begin{aligned} h &: \mathcal{X} \to \mathcal{Y} \\ y &= h(x), \end{aligned} \tag{2.3}$$

which is called *measurement operator*. Given an element of the model space $x \in \mathcal{X}$ one can obtain corresponding data $y \in \mathcal{Y}$ according to this operator.

In general, the operator h is non-linear. In the special case of a linear operator, it is denoted as H and Eq. (2.3) then takes, as usual, the following form:

$$y = Hx. \tag{2.4}$$

Note that methods from topology and especially differential topology [195, 226] exist that would allow us to relax above assumptions. For most of the discussion in this chapter, it would be sufficient when the model and data spaces were differentiable manifolds with a metric. As this generalisation would result in considerable amounts of technical machinery — which would obfuscate the main text while not significantly contributing to the results — it is omitted. The interested reader is referred to the monograph of Lee [226] for further information on smooth manifolds.

2.1.1. Transformations to Vector Space

It is worth pointing out that the seemingly abstract assumption of Banach (Hilbert) vector spaces has very practical consequences. A simple example is $\mathbb{R}^+$, the positive real line, which may be used to describe, for example, a conductivity parameter. This is clearly not a vector space as for $x_1, x_2 \in \mathbb{R}^+$, the result of $x_1 - x_2$ is not necessarily an element of $\mathbb{R}^+$.

[2] See the book of Bourbaki [73] for further information on vector spaces.

The solution is to use an invertible mapping so that the co-domain is a vector space. This could be, in this example, the natural logarithm log together with its inverse, exp, which continuously map $\mathbb{R}^+$ to $\mathbb{R}$ and back. This construction can be generalised to more complex examples, such as symmetric positive definite (SPD) matrices and diffusion tensors. These, as well as a discussion of the relations to Lie groups and Lie algebras [226, p. 93], are given in [271, 230, 30, 31].

In the following, two invertible transformations of common intervals to a vector space are described which can help out.

2.1.1.1. Log-Normal Transform

For a strictly positive scalar quantity $\kappa \in \mathbb{R}_+$ an often used method to transform it to a vector space is to apply a logarithmic transform.

$$\acute{\kappa} = \log_{10}(\kappa) \tag{2.5}$$

The result $\acute{\kappa}$ can be safely used in algebraic manipulations, since $\acute{\kappa} \in \mathbb{R}$, and $\mathbb{R}$ is a vector space. Afterwards, the inverse transform

$$\kappa = 10^{\acute{\kappa}} \tag{2.6}$$

is applied and the scalar is again strictly positive, as required by the application.

2.1.1.2. Inverse Hyperbolic Tangent Transform

For a scalar quantity ϕ lying strictly within a certain open interval $(a, b) \subsetneq \mathbb{R}$ a method to transform it to a vector space is to linearly map the interval (a, b) to $(-1, 1)$ and then to apply the *inverse hyperbolic tangent* function

$$\operatorname{atanh}(x) = \tanh^{-1}(x) := \frac{1}{2} \ln\left(\frac{1+x}{1-x}\right)$$

which maps $(-1, 1)$ to $\mathbb{R}$.

In detail, the mapping consists of the following manipulations:

$$\begin{aligned} \acute{\phi} &= \frac{2\phi - (a+b)}{b-a} \\ \mathring{\phi} &= \operatorname{atanh}\left(\acute{\phi}\right) \end{aligned} \tag{2.7}$$

The result $\mathring{\phi}$ can be safely used in algebraic manipulations, since $\mathring{\phi} \in \mathbb{R}$, and $\mathbb{R}$ is a vector space. Afterwards, the inverse transform

$$\begin{aligned} \acute{\phi} &= \tanh\left(\mathring{\phi}\right) \\ \phi &= \frac{(b-a)\acute{\phi} + (a+b)}{2} \end{aligned} \tag{2.8}$$

is applied and the scalar is again strictly within a certain open interval (a, b), as required by the application. Remember that the *hyperbolic tangent* is defined as

$$\tanh(x) := \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

It is worth mentioning that this transform can also be used to enforce heuristic or physical limits on a quantity. For example, the value for soil porosity lies by definition in the interval $(0, 1)$. However, depending on the specific application, it may make sense to limit the interval to an even smaller interval, *e.g.* $(0, 0.32) \subset (0, 1)$. Then it is *by construction* that, in this example, the upper value of 32% porosity can never be exceeded by any manipulation which is applied to the transformed quantity.

Additional examples of possible transform functions are the *logit* and *probit* functions. They can be applied in an analogous way.

2.1.2. Observations

At this point it is important to mention that there is a 'special kind' of data

$$z \in \mathcal{Y}. \tag{2.9}$$

which is called *observation* and denoted by z. It is the numerical representation of information obtained from the real system of interest, using real physical procedures.

Examples of observations are pressure head data from a well obtained by a pressure sensor, radar measurements of the atmosphere obtained from a radar array, or radiation measurements of a wood obtained by some sensor in a satellite. All these measurements have in common that they are *indirect* in the sense that one has to use a *device* to observe the real physical quantity. Usually the situation is even more complex: one is interested in a physical quantity, but it is possible only to observe a different, related one. This situation is the main reason for introducing separate model and data domains in Eq. (2.1).

In the following it is important to distinguish carefully between *observations* and *simulated data*, since these terms are a constant source of confusion. An observation is data obtained from 'reality', or from a 'simulated reality' in case of perfect model experiments. Simulated data is exclusively simulated using mathematical and numerical procedures. Therefore, especially in the context of dynamical systems, it is also called a *forecast*.

2.1.3. Dynamical Systems

In applications the joint model-data space $\mathcal{M}$ is often — at least partly — time dependent, leading to what is called a *dynamical system*. In accordance with other works the time dependence is denoted by a small index t

$$(\cdot)_t := \mathcal{T} \to \mathcal{M}, \tag{2.10}$$

with $\mathcal{T}$ being the time domain. This index may be attached to all time dependent quantities and operators.

The time dependent part of the model manifold $\mathcal{X}$ is called *state space* (sometimes *phase space*) and denoted by $\mathcal{U}$ whereas the time independent part is called *parameter space* and denoted by $\mathcal{V}$ — so, $\mathcal{X} = \mathcal{U} \times \mathcal{V}$. The elements of the state space are usually just called *state* and denoted by $u \in \mathcal{U}$, the members of the parameter space are called *parameters* and denoted by $v \in \mathcal{V}$, and therefore

$$x_t = (u_t, v). \tag{2.11}$$

Across the literature there is a lack of a common name for the members of the model space. Thus we choose to call them *model quantities* or, for brevity, *model*. For convenience we may write x_t, thus indicating a time

dependence for the whole model quantities — but this means still that, of course, only the state is time dependent.

The behaviour of the dynamical system is described by a so-called *model operator*

$$\tilde{g} : \mathcal{X} \to \mathcal{U} \tag{2.12}$$

that describes the evolution of state u_t in the following way:

$$\frac{\mathrm{d}u_t}{\mathrm{d}t} = \tilde{g}(u_t, v). \tag{2.13}$$

This describes a so-called *Markov process*[3]. In other works, the model quantities and the model operator introduced above may both be denoted as 'model' and we may do so in this work as well — but it will always be clear from the context if the quantities or the operator are meant. To add to the confusion, in other works the operator h is sometimes called 'model' in the case of a non-dynamical system.

This work exclusively deals with time dependent systems. The time index is usually kept for clarity of notation, but it is worth mentioning that considering time independent problems in this framework is straightforward, since they can be seen as a special case of a time dependent system.

2.1.3.1. Discrete-Time Markov Process

Let us consider a dynamical system which is described by the model $x_t = (u_t, v)$. In practice, it is typical to obtain observations from the system at discrete points in time. Measurement devices are usually constructed that way: they need a finite, non-zero amount of time to create a data sample[4]. Thus, it is suitable to consider the state of the dynamical system at discrete points in time,

$$u_{t+\Delta t} = g(u_t, v, \Delta t) := u_t + \int_t^{t+\Delta t} \tilde{g}(u_\tau, v)\, \mathrm{d}\tau, \tag{2.14}$$

[3] See also the monographs by Stroock [334] and Ibe [177, p. 45ff], as well as the book of Jazwinski [187, p. 77ff].

[4] The required amount of time can be stunningly small and the measurement frequency very high when compared to typical system evolution time scales. Then it actually may be beneficial to consider the measurement operation to be a continuous process. See the books of Jazwinski [187] and Bain and Crisan [36] for additional information.

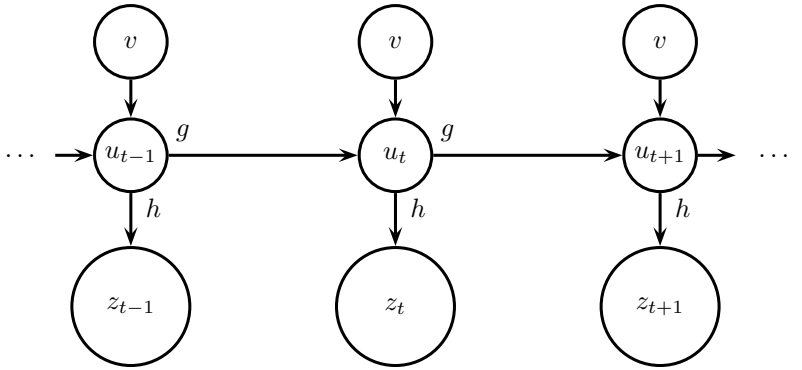

Figure 2.1.: Markov chain model of a dynamical system. The state u_t, parameters v, observation z_t and operators h, g are assumed as perfectly known. Remember that $x_t = (u_t, v)$ — the observation could thus in principle *directly* depend on the parameters, although this is typically the case only indirectly thought the state.

where Δt denotes a positive time increment. Note that it is common to denote the time points by countable indices $\ldots, t-1, t, t+1, t+2, \ldots$, as the actual value of Δt usually does not matter.

It is assumed that observations depends only on the current state, and not previous ones. Thus, the observations z_t, as well as the 'state snapshots' u_t, form a first order *Markov chain*[5] which can be modelled by a *Bayesian network*[6]. A network depicting the complete dynamical system meta model is shown in Fig. 2.1.

2.1.4. Spatial Systems

The systems considered in this work, and thus the spaces $\mathcal{X}$ and $\mathcal{Y}$, usually have a spatial extent. For example, in weather forecasting many of the

[5] See also the book of Whittle [359, chapter 9] for further information on discrete Markov processes.

[6] See also the article of Ghahramani [137] and the book of Jensen and Nielsen [188] for further information.

quantities describing the system are usually defined on a sphere with a certain thickness, modelling the atmosphere of the earth.

By the example of the space $\mathcal{X}$, the spatial extent is introduced via an index set $\mathcal{X}_s$, while $\mathcal{X}_q$ is the usual vector space of physical quantities:

$$\mathcal{X} \subset \mathscr{F}(\mathcal{X}_s; \mathcal{X}_q) = \mathcal{X}_q^{\mathcal{X}_s}. \tag{2.15}$$

$\mathscr{F}(\mathcal{X}_s; \mathcal{X}_q)$ is the set of all maps from $\mathcal{X}_s$ to $\mathcal{X}_q$, and it is a vector space as $\mathcal{X}_q$ is assumed as a vector space (*cf.* [258, p. 28]). In the concrete application a suitable subspace of this space of all functions needs to be chosen, such that the operators g and h are well-defined (*e.g.* the space of n-times continuously differentiable maps, $C^n(\mathcal{X}_s; \mathcal{X}_q)$). This suitable subspace is denoted as $\mathcal{X}$.

Elements from $\mathcal{X}$ are functions of the spatial index x_s, indicated by $x(x_s)$. It is not unusual that the spatial index set has multiple dimensions (see examples below), so that x_s actually is a vector. The notation in this work is, as usual, to denote elements from $\mathcal{X}$ with boldface symbols (*e.g.* $\boldsymbol{x} \in \mathcal{X}$) and omit the explicit spatial dependence. Elements from $\mathcal{X}_q$ are denoted with normal symbols (*e.g.* $x \in \mathcal{X}_q$) and with a slight abuse of notation we may also write $x \in \mathcal{X}$, since it is then clear that a single element from the space of quantities is considered. However, the space $\mathcal{X}_q$ itself may still be quite complex, as a model clearly may consist of several quantities defined over its spatial extent.

A similar construction applies, of course, also to $\mathcal{Y}$. There, the index set $\mathcal{Y}_s$ is often a subset of $\mathcal{X}_s$, depending on the system under consideration.

In the context of geophysical applications, the most common example for $\mathcal{X}_s$ certainly is $\mathcal{X}_s \subset \mathbb{R}^3$, where the $\mathbb{R}^3$ is used to (locally) index the three dimensional structure of the earth's crust. Another, less simple set is $\mathcal{X}_s = S^2 \times [0, a]$, a two-dimensional sphere of thickness a. This could be used to index the the atmosphere of the earth.

2.2. Uncertainties in Models

While creating a numerical model for the system of interest errors are being made which should not be ignored when tackling the task of iden-

tification. Therefore, these errors are discussed in this section[7].

2.2.1. Modelling Errors

To be able to capture a physical, real system into the above framework concrete mathematical structure and abstraction has to be introduced. It is precisely here were so-called *modelling errors* are made. To state just a few, we sometimes are

- purposely neglecting certain aspects of the system which are (1) considered as having negligible effects or (2) being mathematically intractable
- lumping very small scale, very large scale, or computationally otherwise intractable effects into parameters
- making simplifying assumptions on spaces or models with respect to a probably limited ultimate purpose of the mathematical model
- even (inadvertently) choosing inadequate mathematical spaces or models.

Despite these unavoidable errors, this modelling step leads to concrete, mathematically well-defined spaces $\mathcal{X}$ and $\mathcal{Y}$, as well as operators h and g.

The operators h and g are usually modelled by some kind of *partial differential equations* (PDEs), mathematically describing the system behaviour. The spaces $\mathcal{X}$ and $\mathcal{Y}$ are chosen such that the mathematical models used for the operators correctly work.

For example it has been shown that it may be highly important to treat forward modelling errors within data assimilation frameworks for oceanographic systems. And for the example of assimilating satellite radiance data, it may have an important impact to treat modelling errors for the observation operator. Note that unresolved scales in the model some-

[7] Related discussions can be found in *e.g.* Tarantola [340, section 1.3] and Le Maître and Knio [224, section 1.1].

times are separately considered as *representativeness errors*, but here this specific distinction is not made.

2.2.2. Numerical Errors

The mathematical model — specifically the operators Eq. (2.3) and Eq. (2.14) — have to be cast into a computer code. It is this form which is ultimately used in any actual computations. Here, two different sources of errors arise: a discretisation error, resulting from replacing continuous operators in time and space by discrete ones, and a numerical truncation error resulting from the finite numerical resolution of electronic computers[8].

2.2.3. Initial Knowledge Errors

These are errors which arise when trying to determine the initial state and parameters of a numerical model. Obviously, this is only possible up to the 'degree of information' one has. For example, for a subsurface structure in geology, one may possess seismic measurements, bore cores, and the individual knowledge of a geologist combining this information — and still a very significant amount of the structure of the system remains uncertain.

2.2.4. Measurement Errors

When measuring reality, the physical device being used has finite precision. It could be malfunctioning and create false data, or even be operated in an incorrect way. It could measure something one did not expect or it could be disturbed by effects one did not think of and that simply are not part of the model. These issues have to be taken into account when using the data. The usual way to model such errors in the context of dynamic systems is the *hidden Markov chain model* (HMM) (*e.g.* [137, 177, 128, 116]). The measurement error can be a quite important

[8] Let us assume that 'simple' programming errors are not an issue.

quantity and it is easily motivated, so many approaches have been developed to model it. A basic introduction is given by Grabe [145]. An extensive discussion about measurement errors in the context of meteorological models can be found in Lira [234].

2.2.5. Summary

As demonstrated above by enumerating some sources of errors, the numerical model is, for various and quite diverse reasons, not equal to reality. All of these errors should result in a limited degree of trust in the numerical model and its use when controlling a real system. It is obvious that the representation of this limited trust must be an important point in every process.

The basic idea to handle such errors is to, abstractly, put all such 'differences to reality' into the space Ω. Note that even this extended model is, of course, just another mathematical model and thereby still different from reality — but it is *by definition* as close as we can possibly get.

Usually two to three types of error are treated separately because of their quite different structure. For a discussion see *e.g.* Stuart [335, p. 475f]. Another aspect is brought up by Tarantola [340, p. 21], where it is argued that measurement errors and modelling errors usually are within the same order of magnitude, implying that none of them should be ignored. Thus, these different sources of errors are formally put into three different spaces, although they will be lumped together later — but for an entirely separate reason. The space of 'modelling and numerical errors' is denoted by Υ — it is quite common to treat these together[9]. The space of 'measurement errors' is denoted by Γ. And finally, the space of 'initial knowledge errors' is denoted by Θ:

$$\Omega = \Upsilon \times \Gamma \times \Theta. \tag{2.16}$$

We now make the non-trivial assumption that the real system can be represented 'sufficiently close for the application' within the domains of

[9] The numerical error part usually depends on some scaling parameter — for example FEM mesh size — which is assumed to be chosen 'small enough' so that the mathematical modelling error clearly dominates the term.

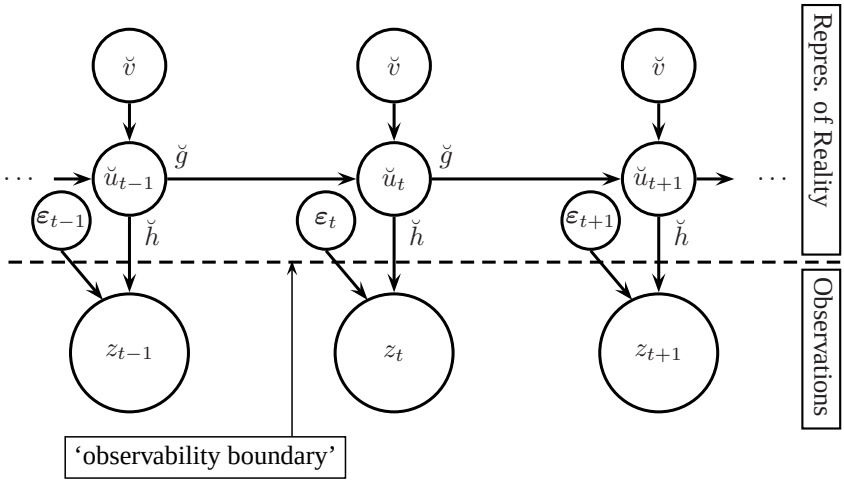

Figure 2.2.: Hidden Markov chain model for the observations. Indirect observations z_t on the assumed 'representation of reality'-model, each disturbed by a random sample $\boldsymbol{\varepsilon}_t = \boldsymbol{\varepsilon}(\omega_t)$, is available. One can think of some kind of observability boundary hiding the 'representation of reality'. Therefore the only possibility is to infer from z_t.

the numerical model, $\mathcal{X}$ and $\mathcal{Y}$. In other words, it is assumed that there exist at any time t elements $\breve{x}_t = (\breve{u}_t, \breve{v})$ and $\breve{y}_t$ which are *representative for reality*. This is certainly a reasonable assumption — otherwise the model should probably be modified — but it should not go by unmentioned. Of course, $\breve{x}_t$ and $\breve{y}_t$ are unknown and the numerical models for g and h contain errors, as discussed above, thereby making it impossible to determine $\breve{x}_{t+1}$ or $\breve{y}_t$ even when given $\breve{x}_t$. The main goal of identification is to derive as much information as possible about $\breve{x}_t$ — including the reliability of this information — from z_t.

When accepting the assumption of a 'representation of reality'-model, one can conceive of a set of maps that — given complete information about all errors — can in some unknown way compute the 'representation of reality'-model, the error-free simulated data, or the observations, respect-

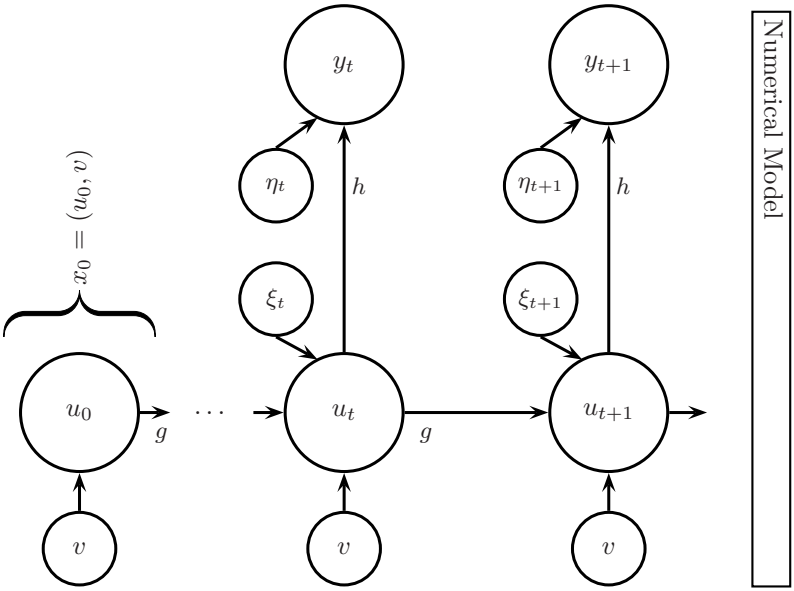

Figure 2.3.: A Bayesian network modelling the relations between quantities of interest in a numerical model of a dynamical system and the involved errors.

ively:

$$x_0 : \Theta \to \mathcal{X}, \tag{2.17}$$
$$\xi_t : \Upsilon \to \mathcal{X}, \tag{2.18}$$
$$\eta_t : \Upsilon \to \mathcal{Y}, \tag{2.19}$$
$$\varepsilon_t : \Gamma \to \mathcal{Y}. \tag{2.20}$$

Here x_0 describes the error in the initial knowledge, ξ_t and η_t describe the errors which arise from mathematical modelling and numerics of operators g and h, and ε_t describes the difference between the theoretical error-free data $\breve{z}_t$ and the actual observation z_t.

The previously defined functions have the properties that

$$\exists \omega \in \Theta : x_0(\omega) = \breve{x}_0 \tag{2.21}$$
$$\forall t \exists \omega_t \in \Upsilon : \xi_t(\omega_t) + g(\breve{x}_{t-1}) = \breve{x}_t =: \breve{g}(\breve{x}_{t-1}) \tag{2.22}$$
$$\forall t \exists \omega_t \in \Upsilon : \eta_t(\omega_t) + h(\breve{x}_t) = \breve{y}_t =: \breve{h}(\breve{x}_t) \tag{2.23}$$
$$\forall t \exists \omega_t \in \Gamma : \varepsilon_t(\omega_t) + z_t = \breve{z}_t. \tag{2.24}$$

In case of observations, the Markov chain model in Fig. 2.1 on page 13 is turned into a HMM, which is modelled by the Bayesian network shown in Fig. 2.2 on page 18. A Bayesian network for the numerical models and the involved errors is shown in Fig. 2.3.

It is obvious that the information on the 'error space' Ω and thus on the error functions, is incomplete. Therefore, the next step is to find a mathematical way to treat this 'incompleteness', thereby enabling its incorporation into practical computations.

2.3. Stochastic Modelling of Uncertainty

The description of the system meta-model pointed at diverse sources of errors. The crucial step is to realize that the errors usually are not *totally* unknown. Typically some incomplete information on the unknown errors is either available, can be assumed, or can be obtained by a method. Thus, limited — possibly subjective — knowledge is available. The question is, of course, how to make use of this knowledge.

The mathematical framework of this work uses a *stochastic* (sometimes also called *probabilistic*) or *information theoretic* approach to the problem. We would like to point out that in our opinion the name 'information theoretic' is much more suitable, but owing to a more conventional nomenclature, we will also use the former one. The approach is inspired by works like Luenberger [245], Halpern [149], Jaynes [186] and Tarantola [340]. In some sense it is an extension to functional analytic approaches given by McLaughlin and Townley [260], where typically a *single* solution to the arising inverse problem is searched for. It can be seen in a broader context as the idea of putting the concept of knowledge and the process of deriving conclusions into a rigorous mathematical framework (*Bayesian brain*; *cf.* [185, 211, 108]).

Stochastic approaches aim at precisely quantifying all possible pieces of information which are available on a certain problem and estimating their influence on the desired goal. Additionally, they use that quantification to improve understanding of the errors and to guide future work being spent on purposefully reducing those errors. This is, or rather should be, just 'good scientific practice' — but there are some important differences:

1. Stochastic approaches allow for incomplete — for example statistical — information. Thus they can be seen as an extension of Aristotelian deductive logic, where only *complete* information is allowed (*cf.* [186, p. 35]).

2. Even more importantly, stochastic approaches put the step of deriving conclusions into a *rigorous mathematical framework* which can be used to actually *perform computations*.

3. Furthermore they can also provide a unifying interpretation of other approaches to the problem, such as regularisation techniques (*e.g.* [335, p. 460]).

2.3.1. Probability Spaces and Random Variables

The unknown errors from Ω cause uncertainty on the joint model-data space $\mathcal{M}$. In section 2.2.5 this causality is represented by the maps defined in Eqs. (2.17)–(2.20). However, to make progress towards practical computations additional structure is necessary.

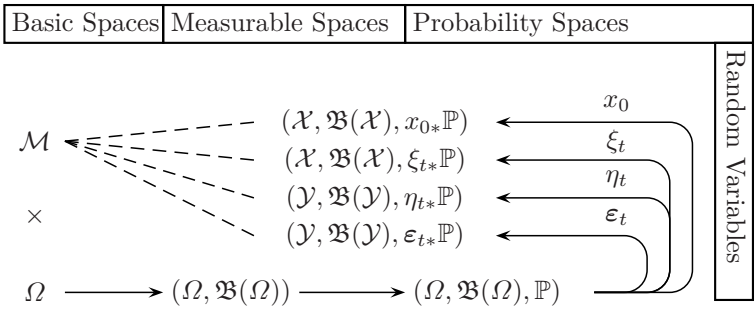

Figure 2.4.: Depiction of the 'pushing-forward' of probability measure $\mathbb{P}$ by RVs x_0, ξ_t, η_t and $\boldsymbol{\varepsilon}_t$ from Ω to the joint model-data domain $\mathcal{M}$.

The intention of probability theory is to formalise statements like "I have the information that $x \in \mathcal{X}$ certainly lies within the interval $[0, 1]$, and it is quite likely, with maybe 70%, that it lies in the interval $[0.3, 1]$. Additionally, I have made the observation that $z \in \mathcal{Y}$ is $0.2 \pm 0.02\%$." Such statements can often be made in practical applications, and they represent the degree of information one has on a problem. The information may come from statistics made from observations or subjective degrees of belief (*e.g.* 'engineering knowledge') — it can all be put into this framework[10].

The central idea of probability theory is to consider measurable maps, so-called *random variables* (RVs)

$$r(\omega) \in L_0(\Omega, \mathfrak{S}, \mathbb{P}; \mathcal{M}, \mathfrak{B}(\mathcal{M})) \tag{2.25}$$

where $\mathfrak{S}$ is a σ-algebra of subsets from Ω and L_0 is the Lebesgue space of $\mathbb{P}$-*measurable* functions w.r.t. the σ-algebras $\mathfrak{S}/\mathfrak{B}(\mathcal{M})$[11]. Many authors limit the term 'RV' to maps with co-domain $\mathbb{R}$, and some thus call above definition *generalised RV* (*cf.* [129]) but here such distinction is not being made. Also note that the same construction applies to $\mathcal{X}$ and $\mathcal{Y}$ as co-domains individually, though here it is presented only for the joint domain $\mathcal{M}$. In cases of ambiguity RVs are distinguished from deterministic quant-

[10] We consider the quasi-religious discussion around subjectivist vs. objectivist interpretation of probability, frequentist vs. Bayesian views etc. as fruitless. The presented approach is mathematically well-founded, which we regard as sufficient.

[11] See [33, p. 36f] and [62, p. 185] for details.

ities by the attached '(ω)', thereby making the parametric dependence on ω explicit.

If the involved σ-algebras and measures are understood, one denotes such functions simply as *measurable*. The elements in $\mathfrak{S}$ are called *events*, and the probability measure $\mathbb{P}$ associates a *probability of occurrence* with each event

$$\mathbb{P} : \mathfrak{S} \to [0, 1], \tag{2.26}$$

$$\forall E \in \mathfrak{S} : \mathbb{P}(E) = p \in [0, 1]. \tag{2.27}$$

The triple $(\Omega, \mathfrak{S}, \mathbb{P})$ is called a *probability space*. Each RV $r(\omega)$ induces a so-called *push-forward* probability measure[12] $r_*\mathbb{P}$, the *probability distribution* or the *distribution measure* of the random variable[13] on the measurable space $(\mathcal{M}, \mathfrak{B}(\mathcal{M}))$ via

$$\forall E \in \mathfrak{B}(\mathcal{M}) : \mu_r(E) := r_*\mathbb{P}(E) := \mathbb{P}(r^{-1}(E)), \tag{2.28}$$

where $\mathfrak{B}(\mathcal{M})$ is the σ-algebra of Borel sets of $\mathcal{M}$. This algebra exists since $\mathcal{M}$ is a vector space and thus a topological space. See for example the classical work of Kolmogorov [212] for further information on probability spaces and events, or the monographs by Ash [33] and Rosenthal [305] for more recent introductions.

Random variables are the basis to treat the elements in $\mathcal{M}$ as stochastic quantities. The maps Eqs. (2.17)–(2.20) are considered as such RVs:

$$x_0, \xi, \in L_0(\Omega, \mathfrak{S}, \mathbb{P}; \mathcal{X}, \mathfrak{B}(\mathcal{X})), \tag{2.29}$$

$$\eta, \varepsilon \in L_0(\Omega, \mathfrak{S}, \mathbb{P}; \mathcal{Y}, \mathfrak{B}(\mathcal{Y})). \tag{2.30}$$

See Fig. 2.4 on page 22 for a graphical representation summarising the whole construction done in this section. RVs will from now on be the main objects of interest, so a closer look at the structure these special functions have and create is worthwhile.

Tarantola [340] and also Stuart [335] take related views on the topic, but they are considering mostly the measure $\mathbb{P}$ and its push-forward $r_*\mathbb{P}$, not

[12] The push-forward is sometimes also called a *transformed measure*, *e.g.* [62, p. 185f], or the *law of the RV*, *e.g.* [94, p. 1].

[13] The construction is not limited to probability measures, but here it is sufficient to consider only those.

directly the RVs pushing them forward. However, interesting results have been derived from this view. For example Cotter et al. [93] combine the stability of inverse problems (described in [335]) with error estimates from the underlying forward problem. The approach taken here is sometimes called the 'Bernoulli' approach [317, p. 233], hinting that it may be considered as an early, quite natural approach to probability theory (*cf.* [317, chapter 8]).

Some important properties of random variables such as the extension to random fields and processes, their connection to information content and probability densities, their algebraic properties, and related important functionals, are summarised in appendix A and used from here on. The reader already familiar with these topics may safely continue with the main text.

2.4. The Stochastic Forward Problem

Having introduced how to mathematically model uncertainty by RVs, an important step towards the ability to perform stochastic *inverse* modelling is to understand how to solve stochastic *forward* problems. This essentially amounts to finding methods which can actually solve the numerical model described in Fig. 2.3 on page 19 when the involved errors and incomplete knowledge are modelled by RVs. Fig. 2.5 depicts a generic stochastic forward problem for the meta model presented in this chapter.

The question is: given descriptions of $x_0(\omega) = (u_0(\omega), v(\omega))$, how to approximate the influence of these uncertain quantities on the future of the dynamical system $x_t(\omega)$ and the predicted measurements $y_t(\omega)$? Or, to put in another way: how to propagate this information through the numerical operators g and h?

The propagation of input uncertainties to output uncertainties is considered in the field of *uncertainty quantification* (UQ). The available methods are tied to the chosen numerical representation of RVs (*cf.* section 3.2) and a suitable subset will be reviewed in section 3.3.

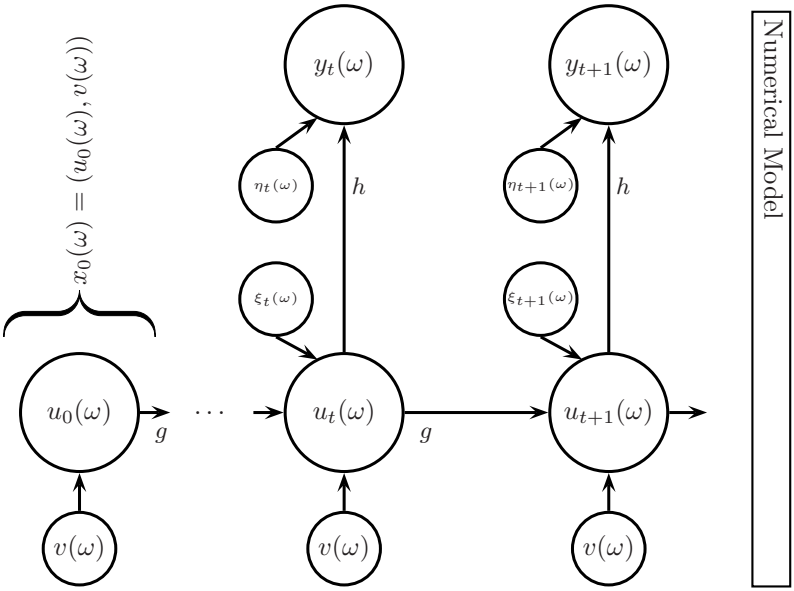

Figure 2.5.: Dynamical stochastic forward problem. The involved uncertain quantities are modelled by RVs, emphasized by their dependency on ω.

2.5. The Stochastic Inverse Problem

The inverse problem is in some sense the opposite to the forward problem discussed in the previous section: the goal is to derive information on the model $x_t(\omega)$ from the observation z_t — the measured 'output of reality'. Note that the assumption of an observation (or measurement) error $\boldsymbol{\varepsilon}_t(\omega)$ turns the observation into an RV, too:

$$z_t(\omega) := z_t + \boldsymbol{\varepsilon}_t(\omega). \tag{2.31}$$

One may also consider multiplicative error terms, but here we limit ourselves to additive ones. In the deterministic setting inverse problems are usually ill-posed[14], as the amount of observations is typically not sufficient to determine the model quantities. For the stochastic setting it is not ill-posed (*cf.* [335]), which can intuitively be understood: missing data and discontinuities are turned into a continuous notion of 'lack of information', and the inverse problem for this representation is well-posed again.

In the following, $(\cdot)_{a:b}$ with $a, b \in \mathbb{N}, a \leq b$, is the shorthand for a sequence $(\cdot)_a, (\cdot)_{a+1}, \ldots, (\cdot)_b$. The solution to the stochastic inverse problem is to combine the information content from the initial knowledge $x_0(\omega)$, the observations $z_{1:t}(\omega)$, the simulated model $x_{1:t}(\omega)$ (computed with the help of g), the predicted measurements $y_{1:t}(\omega) := (h(x_1(\omega)), ..., h(x_t(\omega)))$, the involved modelling errors $\eta_{1:t}(\omega)$ and $\xi_{1:t}(\omega)$, and the measurement error $\boldsymbol{\varepsilon}_{1:t}(\omega)$ into a new RV $\hat{x}_t(\omega)$. This new RV shall represent the combined information about $\breve{x}_t$. In other words, we have

$$\hat{x}_t(\omega) = F(x_{0:t}(\omega), z_{1:t}(\omega), \eta_{1:t}(\omega), \xi_{1:t}(\omega), \boldsymbol{\varepsilon}_{1:t}(\omega)) \tag{2.32}$$

for some mapping F. The result $\hat{x}_t(\omega)$ is called an *estimator* for $\breve{x}_t$. Due to the Markov property of the system under consideration (*cf.* section 2.1.3.1) — and the usual assumption of i.i.d. errors — it is sufficient to consider for each $t > 0$

$$\hat{x}_t(\omega) = F(x_t(\omega), x_{t-1}(\omega), z_t(\omega), \eta_t(\omega), \xi_t(\omega), \boldsymbol{\varepsilon}_t(\omega)). \tag{2.33}$$

Note that in an extended identification setup, we could similarly try to derive estimates for the involved modelling errors $\hat{\xi}(\omega)$ and $\hat{\eta}(\omega)$, and even

[14] Ill-posed in the sense of Hadamard: the solution of an ill-posed problem does not need to exist, when it exists it is usually not unique, and it often also does not continuously depend on the data.

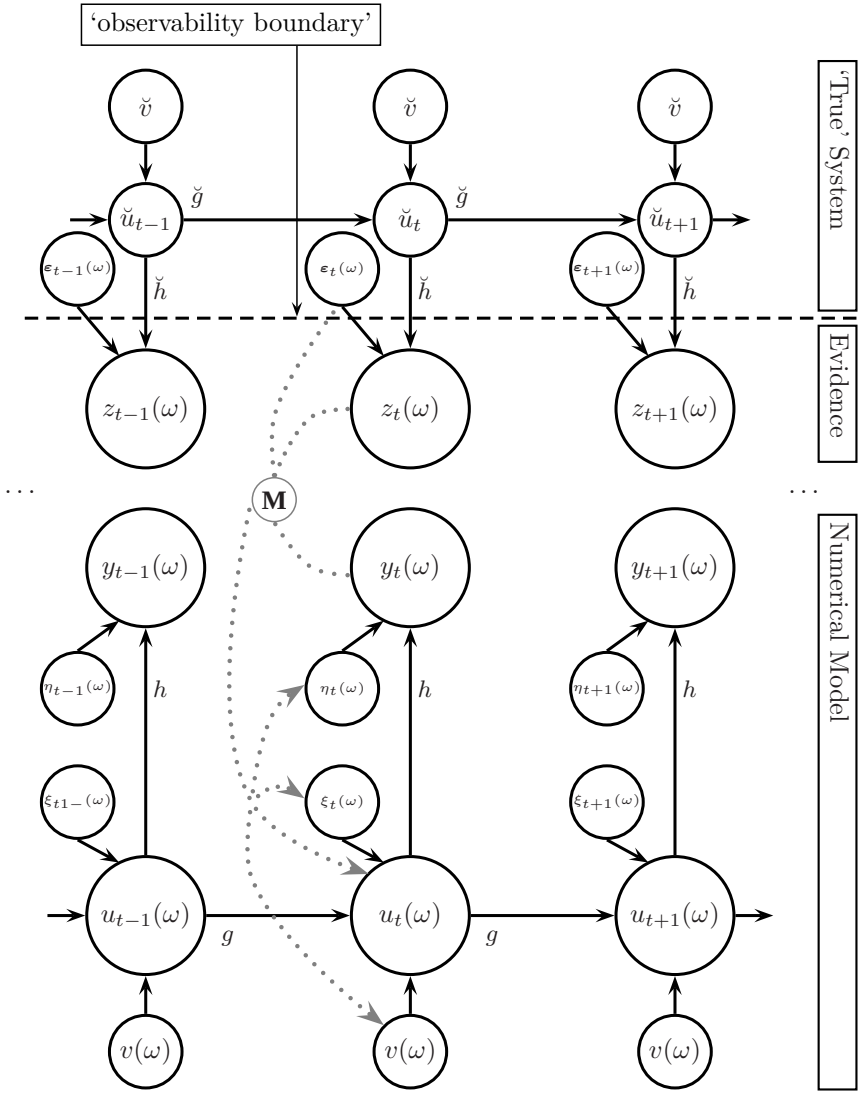

Figure 2.6.: General sequential identification setup. We have noisy observations of a system according to Fig. 2.2 and numerical operators g, h which model this system. We develop identification methods, depicted by **M**, which use the information sources shown by the grey dotted lines in the figure to update the indicated parts of the stochastic numerical model and making it thus a better estimator of $\breve{x}_t$.

the measurement error $\hat{\boldsymbol{\varepsilon}}(\omega)$, from the observations, *e.g.*

$$\hat{\xi}_t(\omega) = \tilde{F}(x_t(\omega), x_{t-1}(\omega), z_t(\omega), \eta_t(\omega), \xi_t(\omega), \boldsymbol{\varepsilon}_t(\omega)). \tag{2.34}$$

However, since this problem has the same structure as the one of Eq. (2.33) it is not considered separately in the following. If any updated information on the errors is available, it is assumed to be used where applicable. If only an a priori estimate is available, this is assumed to be used. If no estimate is available for the modelling errors, typically 0 is used.

This process is usually called *updating*, as we can think of the mapping F as an update to $x_t(\omega)$ with the additional information contained in $z_t(\omega)$, resulting in an improved estimator $\hat{x}_t(\omega)$. In this work, all updated quantities will be denoted by $\hat{\cdot}$, as it is done in most other literature. The updated quantity $\hat{x}_t(\omega)$ should obviously fulfill some optimality criterion, in the sense that $\hat{x}_t(\omega)$ includes as much information as possible — ideally *all* — from the RVs on the right hand side of the equation.

This step is also called *inverse modelling*, *identification*, *model calibration* — depending on the application and methodological context. The generic setup for this step, adopted for the presented mathematical framework, is depicted in Fig. 2.6.

Let us point out that in contrast to classical estimation theory (*e.g.* [227]), the solution of this approach is clearly not a *single*, in some sense best estimator, like *maximum likelihood* (ML) or *maximum a posteriori* (MAP). Such estimators can be *computed from the solution*, though.

2.5.1. Conditional Expectation

To ensure that $\hat{x}_t(\omega)$ includes as much information as possible from the RHS, a suitable mapping F has to be defined. A very specific one is the *conditional expectation* (CE), which is an object of main interest in the present work.

For two RVs $\phi(\omega), \psi(\omega) \in L_1(\Omega; \mathcal{M})$ the information represented by $\psi(\omega)$ can be related to $\phi(\omega)$ using the CE. It intuitively represents the information of what one has to expect from the RV $\phi(\omega)$ given the information $\psi(\omega)$.

The CE of $\phi(\omega)$ given $\psi(\omega)$ is another RV $\hat{\phi}(\omega) \in L_1(\Omega; \mathcal{M})$ for which the probability of every event E in the σ-algebra generated by $\psi(\omega)$ — representing additional information — is equal to the probability of that same event E in $\sigma(\phi)$ (*cf.* [68, definition 3.2.5]):

$$\forall E \in \sigma(\psi) : \int_E \phi(\omega)\, \mathbb{P}(\mathrm{d}\omega) = \int_E \hat{\phi}(\omega)\, \mathbb{P}(\mathrm{d}\omega). \tag{2.35}$$

The RV $\hat{\phi}(\omega)$ is also denoted as the *updated* or *posterior* RV. The usual notation for the CE is

$$\hat{\phi}(\omega) =: \mathbb{E}\left(\phi \,|\, \psi(\omega)\right), \tag{2.36}$$

by which it is made explicit that $\hat{\phi}(\omega)$ is a function of the elementary events ω of $\psi(\omega)$. However, the dependence on ω is often omitted in Eq. (2.36). Note that this dependence is sometimes written like $\hat{\phi}(\omega') = \mathbb{E}\left(\phi \,|\, \psi(\omega = \omega')\right)$ or by abuse of notation like $\hat{\phi}(\omega') = \mathbb{E}\left(\phi \,|\, \psi = \omega'\right)$, especially in the important special case of discrete RVs.

It can be shown via the *Doob-Dynkin* lemma (see [68, p. 90]) or the *Radon-Nikodým theorem* (see [68, p. 103]) that the CE is a function of $\psi(\omega)$,

$$\mathbb{E}\left(\phi \,|\, \psi\right) = f(\psi), \tag{2.37}$$

for a certain measurable function f. For this to hold it is necessary that $\sigma(\psi) \subset \sigma(\phi)$, which is — for example — the case if $\psi(\omega) = \zeta(\phi(\omega))$ for some measurable function $\zeta(\cdot)$. This is obviously the case in the stochastic inverse problem, since in this context $\psi(\omega)$ is an observation, and therefore certainly connected to the representative model $\breve{x}_t$ through the measurement operator h. All that is necessary to assume is that the measurement operator is measurable in the measure theoretic sense[15]. On the other hand, if $\sigma(\psi) \cap \sigma(\phi) = \emptyset$, the two RVs are independent and $\mathbb{E}\left(\phi \,|\, \psi\right) = \mathbb{E}\left(\phi\right)$. Here, the intuitive meaning is that $\psi(\omega)$ did not contain any information which could be related to $\phi(\omega)$, and therefore the expectation of $\phi(\omega)$ did not change.

For further information and proofs see the works of *e.g.* Bobrowski [68, chapter 3], Billingsley [62, p. 445ff], or Bogachev [69, p. 140f]. See the book of Janson [184, p. 127ff] for the special case of Gaussian RVs. The *Fujisaki-Kallianpur-Kunita* equation [130] is a partial differential equation

[15] Please bear with the author on that inevitable sentence.

which describes the time evolution of the conditional expectation and can also be used for the derivation. However, it is not very practical since 'The complicated non-linear structure of this time evolution prevents direct numerical analysis [...]' [244, p. 287].

The equivalence of computing the conditional expectation and computing Bayes formula for conditional probabilities is demonstrated by Bobrowski [68, section 3.2]. However, it is worth mentioning that for $\phi, \psi \in L_1(\Omega, \mathfrak{B}(\Omega), \mathbb{P}; \mathcal{M})$ the conditional expectation $\mathbb{E}(\phi \,|\, \psi(\omega))$ can also be stated as an expectation with respect to a conditional probability measure [70, p. 356ff, esp. proposition 10.4.18]

$$\mathbb{E}(\phi \,|\, \psi(\omega_2)) = \int_\Omega \phi(\omega_1) \mathbb{P}(\mathrm{d}\omega_1 | \psi(\omega_2)). \tag{2.38}$$

In case this conditional probability distribution has a probability density $p(\cdot)$ w.r.t. the probability measure $\mathbb{P}$, we can write this as follows

$$\mathbb{E}(\phi \,|\, \psi(\omega_2)) = \int_\Omega \phi(\omega_1) p(\phi(\omega_1) | \psi(\omega_2)) \mathbb{P}(\mathrm{d}\omega_1) \tag{2.39}$$

and theoretically could use Bayes theorem (*cf.* [42, 222]) to compute $p(\phi|\psi)$. This path quickly leads to the classical approach of treating the densities $p(\cdot)$ and the measures $\mathbb{P}(\cdot)$ as primary objects of interest for probability theory. Therefore let us now look a bit deeper at the conditional expectation itself.

2.5.2. Conditional Expectation in a Hilbert Space of Random Variables

For $\phi, \psi \in L_2(\Omega; \mathcal{M})$ the CE is especially convenient in both interpretation and computation:

$$\zeta(\omega) = \mathbb{E}(\phi \,|\, \psi(\omega)) \tag{2.40}$$

is the *orthogonal projection* of $\phi(\omega)$ onto the subspace

$$L_2(\Omega, \sigma(\psi), \mathbb{P}; \mathcal{M}) \subseteq L_2(\Omega, \mathfrak{B}(\Omega), \mathbb{P}; \mathcal{M}).$$

In other words it is the $L_2(\Omega; \mathcal{M})$-optimal (in the respective inner product; see also Eq. (A.12) in the appendix) estimate of $\phi(\omega)$ in the subspace 'spanned by the information' $\psi(\omega)$.

Mathematically speaking we have for all measurable functions $\vartheta \in L_2(\Omega, \sigma(\psi))$ (*cf.* [359, p. 89, Eq. 14])

$$\langle \phi - \mathbb{E}(\phi \,|\, \psi) \,|\, \vartheta(\psi) \rangle = 0. \tag{2.41}$$

See [152, p. 32–34] or [71, chapter 10.2] for details and further information.

To put it in another way: of all functions of $\psi(\omega)$, $\mathbb{E}(\phi \,|\, \psi)$ is the best predictor of $\phi(\omega)$, in the sense of minimising the *mean square error*. In the special case of centred RVs, it is the *unbiased minimum variance estimate* of $\phi(\omega)$ in the subspace of functions of $\psi(\omega)$. This view is, for example, taken by [245, chap. 4], although the connection to conditional expectation is not made. The fact that the CE is an orthogonal projection will be the main tool used in many concrete methods of chapter 3 to actually perform computations.

2.5.3. Sequential Conditional Expectation

Due to the sequential nature of the observations obtained from a dynamical system one typically has to compute

$$\mathbb{E}(x_t \,|\, x_0, z_{1:t}, \eta_{1:t}, \xi_{1:t}, \varepsilon_{1:t}), \tag{2.42}$$

the conditional expectation of the stochastic model $x_t(\omega)$ given the initial information $x_0(\omega)$ and sequences of observations and errors.

As we assume that our model is a first order Markov process (*cf.* section 2.1.3.1) and that the measurement errors and model errors are uncorrelated in time (*cf.* [299, appendix],[344, section 8.3.2]) we can process the measurements recursively. This is, for example, shown by Evensen [123, section 7.3.2]. Thus, it will be sufficient to consider just a single point in time. It is trivial to see that

$$\hat{x}_0(\omega) = x_0(\omega) = \mathbb{E}(x_0 \,|\, x_0). \tag{2.43}$$

Using the definition, compute $x_t(\omega) = g(x_{t-1}(\omega)) - \xi_t(\omega)$ (*cf.* Eq. (2.14) and Eq. (2.18)), which is the best estimate for $\breve{x}_t$, given all information up to time $t-1$ (note the subtraction of the model error estimate $\xi_t(\omega)$). Now compute

$$\hat{x}_t(\omega) = \mathbb{E}(x_t \,|\, x_t, z_t, \eta_t), \tag{2.44}$$

the CE of the currently best estimate $x_t(\omega)$ (the *prior*), given the information in the prior $x_t(\omega)$, the additional information $z_t(\omega)$, and (if available) the modelling error of the measurement operator, $\eta_t(\omega)$. As the conditional expectation in $L_2(\Omega;\mathcal{M})$ is an orthogonal projection, it is easy to see (*cf.* [245, p. 92f]) that

$$\hat{x}_t(\omega) = x_t(\omega) + \mathbb{E}(x_t \,|\, z_t - (h(x_t) - \eta_t)) \tag{2.45}$$

is the CE in the subspace spanned by the prior and the additional information. One may see that the CE is expressed as a sum of elements from three subspaces, $L_2(\Omega, \sigma(x_t), \mathbb{P}; \mathcal{M})$, $L_2(\Omega, \sigma(z_t), \mathbb{P}; \mathcal{M})$, and $L_2(\Omega, \sigma(\eta_t), \mathbb{P}; \mathcal{M})$. In order to make it a direct sum, one has to 'take out' the information from $z_t(\omega)$ that was already contained in $x_t(\omega)$ — which is precisely $y_t(\omega) = h(x_t(\omega)) - \eta_t(\omega)$. The convergence of this estimate in the mean square sense, and its relation to the all important martingales is nicely shown by Whittle [359, chapter 14.7]. The construction is related to the notion of *innovation* processes, considered as early as Kolmogorov [213].

2.5.4. Linear Conditional Expectation

As it is difficult to project onto the subspace of *all* measurable functions of the observations one has to make simplifying assumptions or allow approximate solutions. The most common ansatz is to assume that the CE is not an *arbitrary* function of $\psi(\omega)$ (*cf.* Eq. (2.37)), but a linear (or rather, an affine) function (*cf.* [18, section 5.2]). For $\phi, \psi \in L_1(\Omega, \mathfrak{B}(\Omega), \mathbb{P}; \mathcal{M})$, set

$$\mathbb{E}\,(\phi|\psi) \approx b + F(\psi(\omega)) \tag{2.46}$$

for some constant b and linear, measurable F. This ansatz we will call the *linear conditional expectation* (LCE) and denote it by

$$\mathbb{E}_{\text{lin}}\,(\phi|\psi) = b + F(\psi(\omega)). \tag{2.47}$$

It is by construction the linear estimator with minimum variance, also called the linear least squares approximation.

It can be shown that the LCE can be computed (*cf.* [245, p. 87], [359,

chapter 14]) by

$$\begin{aligned}\mathbb{E}_{\text{lin}}(\phi|\psi) &= \mathbb{E}(\phi) + \boldsymbol{C}_{\phi,\psi}\boldsymbol{C}_{\psi}^{-1}(\psi - \mathbb{E}(\psi)) \\ &= \underbrace{\mathbb{E}(\phi) - \boldsymbol{C}_{\phi,\psi}\boldsymbol{C}_{\psi}^{-1}\mathbb{E}(\psi)}_{=b} + \underbrace{\boldsymbol{C}_{\phi,\psi}\boldsymbol{C}_{\psi}^{-1}}_{=F}\psi(\omega)\end{aligned} \tag{2.48}$$

Comparing this to the 'full' CE, we limit the projection of $\phi(\omega)$ onto the subspace of *linear* measurable functions of $\psi(\omega)$ — a subspace of *all* measurable functions. This space is smaller and its orthogonal complement — which is where the error 'lives' in — is larger. Therefore it must be noted that the information content of the observation is — in the general case — not *fully* used. This is the approximation that LCE approaches make.

In some cases though the LCE directly coincides with the 'full' CE (especially when the involved RVs are jointly normal, *cf.* [359, section 14.3]). But note that LCE is *always* formally equal to the CE if $\psi(\omega)$ is decomposed into a suitable class of 'basis functions' (possibly infinitely many; for example the indicator functions of sets), and then the LCE is performed on those (see [359, section 14.7]). This approach is related to the linear Bayesian regression of Hartigan [157] and the 'Bayes linear' ideas of Goldstein and Wooff [141].

2.5.5. Recursive Linear Conditional Expectation

Under the condition that $\boldsymbol{C}_{x_t z_t} = 0$, the recursive CE construction of Eq. (2.45) directly carries over to the LCE and can be computed as (*cf.* [245, p. 93, Example 1])

$$\hat{x}_t(\omega) = x_t(\omega) + \boldsymbol{K}(z_t(\omega) - (h(x_t(\omega)) - \eta_t(\omega))). \tag{2.49}$$

with

$$\boldsymbol{K} := \boldsymbol{C}_{x_t y_t}\left(\boldsymbol{C}_{z_t} + \boldsymbol{C}_{y_t} + \boldsymbol{C}_{\eta_t}\right)^{-1} \tag{2.50}$$

There, the linear weighting operator computed from the different covariances is typically abbreviated as $\boldsymbol{K}$ and denoted the *Kalman gain*. To retain equality to the CE, the involved RVs have to be jointly normal. Thus it is clear that both the model operator g and the measurement operator h have to be linear, so that the involved RVs *stay* jointly normal

for all time while being propagated through those operators. However, the approach is clearly applicable to any suitable RVs and operators (also non-Gaussian RVs and non-linear operators), but the linear approximation may result, depending on these components, in noticeable errors.

2.6. Outlook: Optimal Control Under Uncertainty

The main topic of this text is identification. However, this is not performed without intent on how to use the results. Therefore, in the following a short overview of a possible extension of above mathematical framework towards control applications is given.

Taking into account all prior information and observations up to time T, the RV $\hat{x}_T(\omega)$ represents the optimal combination of all available information on $\breve{x}_T$. In the language of statistics, this representation is equivalent to a *sufficient statistic* (*cf.* [196, p. 106]) for the incomplete information on the system, meaning that 'no other statistic which can be calculated from the same sample provides any additional information as to the value of the parameter' [127, p. 310]. This RV is the starting point for forecasting the behaviour of the system under consideration — within the limits given by the available information — by solving the stochastic forward problem (*cf.* section 2.4) with the help of the model operator g. A forecast of the future output of the system can be computed with the help of the measurement operator h. At this point it needs to be stressed that these forecasts naturally contain a quantification of the uncertainty on the future system state and performance based on the available information, $\hat{x}_T(\omega)$.

Additionally, the stochastic forward model can be used to make predictions about how *control actions* — denoted by a_t, $t \in [T, T+1, ...]$ — performed on the system will affect its behaviour: the effect of those controls is predicted by simply applying them to the model[16]. It is then quite straightforward to set up an optimisation problem that determines a set of controls that results in a somehow *optimal* future behaviour of the model.

[16] Of course this may be far from simple — but it is a modelling problem which is not the key interest of this work.

Obviously, it is assumed that the optimal controls for the model perform equally well when applied to the real system of interest. Again it is worth pointing out that the model is a stochastic one and that any optimisation approach has to take this into account. Since this implicitly means that all available information on the problem has been systematically included into the solution, our hope is not ungrounded.

This kind of optimisation setup is often termed *decision support system* or *optimal control system* (considered in *optimal control theory*[17]), depending on the application context and the type of decisions which are to be taken. Historically, such problems are solved by *dynamic programming*[18]. The optimisation problem is also known as *reinforcement learning*[19], a branch of *machine learning* which deals with how an abstract agent should take actions within some environment to maximise cumulative reward. This problem is quite general and thus has been dealt with in many scientific disciplines, including robotics, car, and aeroplane navigation.

2.7. Discussion

From a computational point of view, the mathematical framework described in this chapter may be quite demanding. For the actual computations approximations have to be carefully introduced. However, we believe that it is indispensable to state the framework in all its generality. This shows what is, in our opinion, *the* solution to the stochastic identification problem. Starting from this general ground, one may honestly introduce and correctly judge approximations and simplifications, since now it is possible to point out the differences they may create to *the* solution.

This mathematical framework now serves as the basis to review and discuss suitable methods for stochastic identification in a *unified* context in the following chapter 3.

[17] See the monograph of Bertsekas and Shreve [56].
[18] See the works of Bellman [45], Bertsekas [55], and Powell [296].
[19] For further information see for example the work of Heidrich-Meisner and Igel [162] and the references therein.

Chapter 3.

Review of Methods for Bayesian Inference

This chapter discusses established methods to numerically represent random variables and to solve the problem of Bayesian inference. The first section is devoted to preliminary considerations for creating the prior model $\boldsymbol{x}_0(\omega)$ and preparing the observations, because of their practical significance.

Setting up the prior stochastic model plays a central role in the Bayesian framework. As Stuart [335, p. 464] illustrates, for the under-determined identification problem (where data is sparse or has large errors) a significant amount of information from the prior will be carried over to the posterior. As an example, this has been demonstrated several times in the case of history matching of hydrocarbon reservoirs (*cf.* [284, p. 189]). A prior has to be representative or — informally speaking — span the space in which the assumed representative model exists, as stated in the assumption Eq. (2.21). This is especially true in the context of Monte Carlo sampling based inversion methods with small ensemble sizes (*cf.* [281, p. 34]). In other words: the prior has to accurately quantify the uncertainty one has about the model, and it must not be completely wrong in the mean sense.

3.1. Preliminary Considerations Before Setting Up the Prior

Methods for setting up the prior need to provide a description of the incomplete prior knowledge $\boldsymbol{x}_0(\omega)$ that fits the subsequent identification methods (*cf.* section 3.2). However, there are some steps to be performed before actually creating a numerical representation for a prior.

3.1.1. Re-Parametrisation

Often the dimensionality of the stochastic joint model-observation space $\mathcal{M}$ is prohibitively large, leading to computational infeasibility for both storage and computational complexity reasons. Additionally, depending on the parametrisation of the problem, the space may have a complex structure with multi-modal, strange distributions on the RVs, leading to problems in the identification step and the necessity to use high-order methods. Thus it is often advisable to perform a re-parametrisation of the joint model-observation space. This can be done either manually — based on prior knowledge — or automatically and possibly adaptively — based on some suitable optimality criterion.

The complexity reduction often involves manual or half-automatic approaches, as one somehow has to introduce prior knowledge. A simple approach is certainly the normal-score transform, which is for example applied by Zhou et al. [373]. Examples include the re-parametrisation of water saturation in a hydrocarbon reservoir by a 'waterfront arrival time' (*cf.* [88]) and (non-Euclidean) distance-based approaches (*cf.* [252]). Additional approaches are pilot point strategies (*e.g.* [16]), gradual deformation (*e.g.* [174]) and also adaptive techniques like presented by Da Veiga and Gervais [95]. This and additional methods in the context of history matching hydrocarbon reservoirs are discussed at length by Oliver and Chen [284, section 2]. A method which is applicable to higher-order statistics is explained in the work of Jafarpour and Khodabakhshi [181]. Examples in other areas of application exist but are omitted for brevity.

The dimensionality reduction is often performed by automatic approaches and is typically considered as an integral part of solving the stochastic

forward problem, so we discuss them in section 3.3.

3.1.2. Screening of Evidence

An often overlooked but highly important step in the creation of a prior and the setup of any inversion algorithm in a practical application is to screen the observations for common data errors and statistical problems. It may even be the most time consuming part of the whole workflow:

> Careful consideration of these issues is time consuming and sometimes tedious; it is common, for instance, to spend many days in careful examination of data prior to running the main analysis that, itself, takes about 5 minutes. [338, p. 60]

To name just a few issues, one may face: missing observations, outliers; un-normality; heteroscedasticity; heterogeneity of variance and heterogeneity of variance-covariance matrices; multicollinearity and singularity. See the book of Tabachnick and Fidell [338, chapter 4] for a discussion of this topic, and a practical 'Checklist for Screening Data' [338, Table 4.4].

Screening may either be performed semi-automatically with the help of statistical software (*cf.* [338]), or automatically. An example of an automated procedure for data quality control with application to meteorology is presented by Steinacker et al. [329]. The very broadly laid out upcoming family of international standards for industrial data quality will be ISO 8000 (see *e.g.* [49, 357] for short overviews).

Screening of data is not a central topic of the present text, since all numerical applications (see chapter 4.5) are conducted in perfect model scenarios and data quality is therefore ensured. However, it is stressed that in any real application this step may be of central importance.

3.2. Numerical Representation of Random Variables

In the previous chapter, RVs have been introduced as abstract functions on probability spaces. Yet, typically the elements of the sample space, *i.e.* $\omega \in \Omega$, are usually intractable. This leads to the question how to actually *represent* an RV in practical computations.

The representation of RVs is often strikingly different from what is used for 'normal' functions. It usually takes one of the following forms:

Spectral: In these functional approximation methods, the idea is to describe an RV $r(\omega)$ as a function of other RVs of some known, simple type. This has a distinctly functional analytic flavour (*cf.* [317]).

Sampling: Sampling is an evaluation of the RV $r(\omega)$ at some points, randomly or deterministically chosen according to the measure $\mathbb{P}$.

Distribution: This is the push-forward measure $r_*\mathbb{P}$ generated by an RV $r(\omega)$. This description leads to the formulation of conservation equations for this probability, variously known as *Kolmogorov-equations*, *Fokker-Planck-equations*, or *master equations*. For larger models these methods are usually not even contemplated for practical use due to their computational demand (*cf.* [323]).

(Central) Moments of r: These are the quantities

$$M_r^{(k)} := \mathbb{E}\left((r - \mathbb{E}(r))^k\right).$$

This approach leads to — usually ever more complicated — evolutionary integro-differential equations for the moments (*cf.* [323]).

Most contemporary methods rely on the sampling approach (either deterministic or random), but spectral approaches are gaining increased attention (*cf.* [224, section 1.4]). We further consider these two representations and not the others, because of their sometimes strong limitations (*cf.* [363, p. 3f]). However, we may make connections to them where appropriate.

3.2.1. Spectral Representation

Spectral approaches employ convergent series expansions for the involved RVs, *i.e.*

$$r(\omega) = \sum_{\alpha \in \mathcal{A}} r^\alpha f_\alpha(\varphi_1(\omega), ..., \varphi_k(\omega), ...) \tag{3.1}$$

with r^α coefficients, f_α some special functions, $\varphi_i(\omega)$ RVs of a simple, known type, and $\mathcal{A}$ a multi-index set discriminating the functions f_α. Depending on the choice of f_α and $\varphi_i(\omega)$, one ends up with different methods.

Typical examples are the well-known *Fourier expansion* (where the f_α are sine/cosine waves), the *Karhunen-Loève expansion* (KLE; also called principal component analysis, a generalisation of the Fourier approach; *cf.* [15],[138, chapter 2.3]), Wiener's *polynomial chaos expansion* (PCE; also known as *white noise analysis* — there the f_α are multivariate Hermite polynomials, *cf.* [360, 168, 184, 247, 164, 163]), and a generalisation of this approach to other polynomials obtained by the Askey scheme (gPCE; *cf.* [363], [117, section 3]).

By the example of the PCE the ideas, advantages, and disadvantages of such spectral approaches are now discussed. The PCE will be also used later on in this text. Some important aspects of it — which are, however, not necessary for immediate understanding — are covered in appendix B. For the other approaches, the reader is kindly referred to the references given above.

3.2.1.1. A Complete Orthonormal System for the Hilbert Space of Random Variables

A convenient property of Hilbert spaces is that the notion of an orthonormal basis generalises to them. For example it can be shown that for uncorrelated Gaussian RVs $\theta_1, ..., \theta_k, ... \in L_2(\Omega; \mathbb{R})$, arbitrary sums and products (polynomials) exist, since Gaussian RVs have moments of all orders. Additionally, these polynomials are dense in $L_2(\Omega; \mathbb{R})$, thus opening a way to represent any RV from $L_2(\Omega; \mathbb{R})$ arbitrarily well by an infinite series of polynomials of Gaussian RVs. This can be extended to random

fields from $L_2(\Omega;\mathcal{M})$ by equipping it with a Gaussian product probability measure (see appendix B.2). If we now choose multivariate Hermite polynomials in uncorrelated (independent) Gaussian RVs, they create a very special system of orthogonal subspaces, whose union (sometimes called the *Cameron-Martin-space*, *cf.* [82, 83]) is dense in $L_2(\Omega;\mathcal{M})$ as required. See for example the works of Ernst et al. [117, section 2] or da Prato [94, p. 125ff] for an introduction.

We assume that any RF $\boldsymbol{r}(\omega) \in L_2(\Omega;\mathcal{M})$ has an expansion in Hermite polynomials — this is called the *polynomial chaos expansion* (PCE):

$$\boldsymbol{r}(\omega) = \sum_{\alpha\in\mathcal{J}} \boldsymbol{r}^\alpha H_\alpha(\theta_1(\omega), ..., \theta_k(\omega), ...), \tag{3.2}$$

with H_α multi-variate Hermite polynomials (see appendix B.2). The sequence of coefficients $(\boldsymbol{r}^\alpha)_{\alpha\in\mathcal{J}}$ is also called the *Hermite transform* $\mathscr{H}(\boldsymbol{r})$ of the RF $\boldsymbol{r}(\omega)$ (see appendix B.4 and [256]). This sequence represents the RF and may be computed simply by projection:

$$\forall\alpha \in \mathcal{J}: \quad \boldsymbol{r}^\alpha = \mathbb{E}(\boldsymbol{r}(\cdot)H_\alpha(\cdot))/\langle H_\alpha|H_\alpha\rangle. \tag{3.3}$$

RFs form an integration algebra — an algebra with RFs as primitive objects. For the PCE, the implementation of this algebra is described in appendix B.3 and — including certain non-polynomial functions — is also discussed in the works of Debusschere et al. [98, section 2] and Le Maître and Knio [224, section 4.5].

3.2.1.2. Truncation

For the numerical implementation the Hermite transform obviously has to be limited to a finite amount N of Gaussian RVs $\theta_i(\omega)$ and limited to a finite polynomial expansion — in other words to a finite subset $\mathcal{J}_Z$ of the multi-indices $\mathcal{J}$:

$$\boldsymbol{r}_Z(\omega) = \sum_{\alpha\in\mathcal{J}_Z} \boldsymbol{r}^\alpha H_\alpha(\theta_1(\omega), ..., \theta_N(\omega)). \tag{3.4}$$

The simplest way to limit the index set is to truncate the series at a certain highest polynomial degree P. The necessary amount of basis RVs is either determined by the model under consideration, or it is reduced via

problem dependent, truncated spectral expansions such as the truncated KLE.

Recently popular, though, are adaptive choices like the *generalized spectral decomposition* [278], the *stochastic model reduction algorithm* (*cf.* [111]) and the *VLR-SR1U* approach [217]. These approaches adapt the spectral representation basis to the problem, thus making high dimensional applications feasible.

3.2.2. Sampling Representation

For the sample representation, the RV $r(\omega)$ is evaluated at some randomly or deterministically chosen points $\omega_s \in \Omega$. The points have to be chosen according to the probability measure $\mathbb{P}$. This is conceptually the simplest approach as it only needs — usually very many — evaluations of the deterministic model. However, this makes these methods computationally very costly, especially for real applications (*e.g.* [325]).

3.2.3. Computing the Representation

For example for earth science applications it is the area of geo-statistics (*e.g.* [103, 133]) which provides the required representations. Here, one works mainly with random fields (*cf.* appendix A.1). Such RFs are typically defined on $\mathcal{X}_s \subseteq \mathbb{R}^3$ with Cartesian coordinates. Sometimes simplifications to $\mathbb{R}^2$ or even $\mathbb{R}^1$ are made. In subsurface applications, the RFs typically describe parameters (like permeability, porosity or net-to-gross-ratio) or variables (like pressure or fluid saturations). In atmospheric applications, they are usually transient fields like pressure, temperature, or velocity.

Spectral expansions for RFs may be computed from a given or assumed covariance function by solving an eigenvalue problem. The procedure is described, for example, in [256]. Adaptive constructions are more involved and tightly coupled to the forward problem. Thus we refer the reader to above references. To the best knowledge of the author, higher-order statistics are usually not considered for the direct construction of spectral representations.

For the sampling representation, common second order — or variogram-based — algorithms include variants of Kriging, and (correlated) sequential Gaussian simulation (*e.g.* [102, 300, 114]). Efficient methods to create realisations conditioned to variograms and prior data include the work of Reynolds et al. [301]. Higher order geo-statistics is commonly referred to as multiple-point geo-statistics. There, databases of training images play a central role (*e.g.* [333, 332, 81]). All these algorithms can provide random samples of an underlying geo-statistical model and thus correspond to the 'probabilistic sample representation' of an RV.

For further reading on multivariate analysis, a highly recommended reference is the monograh of Izenman [180]. In other areas of application, methods similar to the ones discussed here are used. Many of these methods, especially connected to heterogeneous materials, are discussed by Torquato [342].

3.3. Methods for Solving the Stochastic Forward Problem

Methods for the solution of the stochastic forward problem are tightly coupled to the chosen representation.

Starting with the book of Ghanem and Spanos [138], spectral methods for the solution of stochastic forward problems have become increasingly popular, and a large amount of different methods have been developed [364, 202, 203, 201, 368, 278, 217, 236, 235]. The stochastic forward problem is numerically solved by such spectral methods basically by projecting the constitutive equations onto the spectral representation given by Eq. (3.1). Due to the similarity to the *finite element method* (FEM; *e.g.* [280]) — in fact it can be readily combined — this method is also called the *stochastic Galerkin* approach. A review of such methods can be found in the work of Jakeman and Roberts [182], and a broader overview of stochastic Galerkin methods is given by Matthies [256, 257]. A state-of-the art report of fast numerical methods for stochastic computations is presented by Xiu [362], whereas a short overview of the field is given by Matthies [259]. Recent monographs on the topic include the works of Xiu [363] (general), Le Maître and Knio [224] (with applications to

computational fluid dynamics) and Kopriva [214] (with algorithms and implementation suggestions).

The sampling representations (both deterministic and probabilistic) of RVs simply employ the original deterministic forward model for each of the samples in an 'embarrassingly parallel' way, thereby solving the stochastic one. Approaches to speed up these many, similar computations include the work of Parks et al. [287]. Numerical solutions for the deterministic forward model are, depending on the model, obtained by discretisation approaches for PDEs, like the *finite element method* FEM and related approaches. Deterministic samples usually are constructed from a spectral representation. This approach is denoted as *stochastic collocation*. A review of such methods is given by Jakeman and Roberts [182].

3.4. Methods for Solving the Stochastic Inverse Problem

In this section methods for solving the stochastic inverse problem are discussed. The focus is — due to the dynamic nature of the considered systems — on sequential or recursive methods. Such methods exploit the Markov property of the system meta-model and are also called *filters*. Popular related methods that do not usually exploit the Markov property are discussed shortly in section 3.4.4.

The methods are classified according to the following two categories: *approximation* (linear, approximate non-linear or fully non-linear) and *representation* (deterministic sampling, probabilistic sampling or spectral). The first category distinguishes if and how the method approximates the solution of the conditional expectation (*cf.* Eq. (2.36)). The second category distinguishes how the involved RVs are represented during the update step (*cf.* section 3.2).

This classification allows us to present similarities and differences in a structured way and effectively discuss the advantages and disadvantages of each method. Similar attempts to categorise and review inversion methods are made by *e.g.* Lakshmivarahan and Stensrud [221] (from a meteorological application perspective), Daum [96] (from a real time application

practitioner's perspective) and Wikle and Berliner [361] (from a generic Bayesian perspective). A classical review paper on the topic, which distinguishes linear and non-linear filtering is given by Bucy [76].

3.4.1. Linear Methods

Linear methods solve (or approximate) the recursive LCE Eq. (2.49) in a way tailored to the representation of the involved RVs. Note that this may include methods which allow for non-linear forward and measurement operators — only the estimator is linear.

The limitations resulting from this assumption are shown in a practical example by *e.g.* Posselt and Bishop [294]. There, a stochastic LCE implementation is compared to a stochastic CE update. The authors demonstrate that the LCE has difficulties with the *transition* from an unimodal prior and data to a multimodal posterior distribution. Once the multimodality is established in both prior and posterior, the LCE method is able to update the multimodal distribution. This ability is also demonstrated by Pajonk et al. [8, section 7.1] for a deterministic LCE implementation (*cf.* chapter 4).

A natural problem of all linear methods is the underestimation of the 'true' variance due to limiting the projection to linear functions of the observations (*cf.* section 2.5.4). For example it is demonstrated by Law and Stuart [223] on a 2D Navier-Stokes model in different regimes that the variance estimates from such methods have to be treated with care. However, linear methods are highly popular in practice due to their manageable computational demands.

3.4.1.1. Spectral Representation

The most famous instance of spectral methods for recursive linear conditional expectations is the original *Kalman filter* (KF; [198]). It assumes purely multivariate Gaussian RVs, a linear model operator $g = \boldsymbol{G}$, and a linear measurement operator $h = \boldsymbol{H}$. These restrictions ensure that the model RV $\boldsymbol{x}_t(\omega)$ will always be multivariate Gaussian. In that sense, the Kalman filter is exact and does not have the problem of underestimating

the 'true' variance. The stochastic spectrum of a multivariate Gaussian RV is determined by its mean and covariance structure, which is exactly the representation used in the Kalman filter. It is well known that the posterior RV $\hat{\boldsymbol{x}}_t(\omega)$ updated according to Eq. (2.49) is also multivariate Gaussian and its mean and covariance can be computed according to

$$\hat{\bar{\boldsymbol{x}}}_t = \bar{\boldsymbol{x}}_t + \boldsymbol{K}(\bar{\boldsymbol{z}} - \bar{\boldsymbol{y}}), \tag{3.5}$$

$$\boldsymbol{C}_{\hat{\boldsymbol{x}}_t} = (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H})\,\boldsymbol{C}_{\boldsymbol{x}_t}. \tag{3.6}$$

In this special case the conditional expectation (*cf.* Eq. (2.36)) is linear and its mean coincides with the maximum likelihood solution, as well as the recursive least squares solution (see [187, p. 201ff, examples 7.1–7.3] for a discussion and details).

When the involved operators are non-linear, an approximation called 'the' *extended Kalman filter* (EKF) is usual [187, section 8.3]. There, the operators are linearised with a first order truncation of a Taylor series expansion. This clearly introduces closure problems when the higher order terms in the expansion cannot be neglected (*cf.* [119, 14, 124]). Another major problem of the Kalman filter appears when the model space becomes high dimensional: storage requirements of the covariance matrix $\boldsymbol{C}_{\boldsymbol{x}_t}$ become an issue (*cf.* [14]). We note, though, that 'the' EKF does not exist and is more a family of filters with long lists of engineering tricks (*cf.* [96, p. 59]). A nice introduction to the Kalman filter is [356]. A plain extended Kalman filter is discussed in [187], as well as the time-continuous *Kalman-Bucy* filter. A recent monograph on the topic is [36]. An important implementation variant for the Kalman filter is the *square root* filter [295, 61], which has been used for the Apollo space program. Such a square root implementation is also the central topic of chapter 4.

The singular evolutive extended Kalman filter (SEEK, introduced in [291, 293, 292]) involves a deterministic low-rank representation of the involved covariance matrix, given by a truncated eigen-decomposition. This is inserted into the extended Kalman filter equations, leading to reduced storage costs. An application example in ocean dynamics is given in the work of Rozier et al. [306].

An extension of the linear conditional expectation approach to non-Gaussian RVs represented by spectral decompositions is extensively discussed in chapter 4. Related approaches are [307, 67, 66, 369], which are developed as extensions of the (extended) Kalman filter theory. On the

other hand, the approach discussed in chapter 4 is obtained by a direct projection of the linear conditional expectation onto a spectral representation. Additionally, its implementation is related to the square root filter, which is demonstrated to have some benefits.

3.4.1.2. Deterministic Sampling Representation

The most famous approach for recursive, linear conditional expectations involving deterministic sampling is certainly the class of *sigma-point* or *unscented Kalman filters* (UKF) [194, 191, 190, 192, 193, 345]. These are deterministically chosen sample points which are used to propagate the mean and covariance structure of a multivariate Gaussian RV through any non-linear forward and measurement operators accurately to the 3rd order (Taylor series expansion), thus having less closure problems when compared to the extended Kalman filter (*cf.* [353]). Ito and Xiong [179] describe a method which selects sampling points according to the Gauss-Hermite quadrature rule (so it is similar to the UKF [96, p. 60]), as well as a finite difference approximation. In [29] so-called *cubature Kalman filters* (CKF) are introduced which employ special integration methods to improve accuracy, while the amount of necessary integration points only scales linearly with the dimension of the problem.

The singular evolutive interpolated Kalman filter (SEIK; [291]) is an approximation to the SEEK filter, which involves interpolating states chosen deterministically according to an *empirical orthogonal functions* (EOFs) or *principal components analysis* (PCA)[1]. Thereby the rank of the involved covariance matrix is reduced, but also storage requirements drop.

3.4.1.3. Probabilistic Sampling Representation

The field of *probabilistic* LCE updating methods has certainly received considerable attention since the introduction of the first variant, the *ensemble Kalman filter* (EnKF), by Evensen [120] in 1994. These methods avoid the two major problems of the original Kalman filter (closure problems for non-linear model and measurement operators; prohibitive storage requirements of the involved covariance matrix). Many different, related

[1] Here we can see a clear connection between sampling and spectral representations.

algorithms have been conceived. They are all similar in that they approximately solve the recursive LCE by representing the involved RVs by a Monte Carlo sample called *ensemble*. It has been shown that probabilistic LCE updating methods are applicable to quite high dimensional problems with relatively few ensemble members. Often though, additional regularisation methods have to be included to prevent filter divergence. This frequently results from the Monte Carlo sampling errors introduced by the RV representation. Additionally, variance and higher order estimates should be treated with care, due to the comparably slow convergence of Monte Carlo sampling.

Recent generic overviews of the EnKF and related filters are contained in Biegler et al. [60, chapter 11] and the article of Evensen [124] (which is part of a special issue of the *IEEE Control Systems* journal [221, 248, 25]). A monograh on EnKF and connected topics is given by Evensen [123]. Short reports on the practical implementation of the EnKF are the ones of Mandel [249, 250]. Aanonsen et al. [14] have written a review article on EnKF for reservoir engineering applications. A special collection of articles of the *Monthly Weather Review* related to the EnKF in atmospheric data assimilation is available [266]. A special issue of *Computational Geosciences* on EnKF-related topics is also available [279].

Overview of the EnKF The following presentation is valid for a single time point t, which is therefore omitted from the notation. The ensemble size is $N \in \mathbb{N}$. Each ensemble is conveniently written in matrix form, with one sample in each column: $\boldsymbol{X} = [\boldsymbol{x}(\omega_1), ..., \boldsymbol{x}(\omega_N)]$, and similarly for the forecasted measurements $\boldsymbol{y}(\omega)$ and the observations $\boldsymbol{z}(\omega)$.

All necessary covariances to compute the Kalman update can be approximated from the prior ensemble. For this first centralize $\boldsymbol{X}$ and $\boldsymbol{Y}$:

$$\tilde{\boldsymbol{X}} := \boldsymbol{X} - \bar{\boldsymbol{x}}\mathbf{1}_N^T \tag{3.7}$$

$$\tilde{\boldsymbol{Y}} := \boldsymbol{Y} - \bar{\boldsymbol{y}}\,\mathbf{1}_N^T. \tag{3.8}$$

Here $\bar{\boldsymbol{x}} := \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}(\omega_i)$ denotes the sample mean, and $\mathbf{1}_N^T$ represents a row vector of ones of length N. Then one may estimate the (cross-)

covariances:

$$\boldsymbol{C_{xy}} \approx \frac{\tilde{\boldsymbol{X}}\tilde{\boldsymbol{Y}}^T}{N-1} \tag{3.9}$$

$$\boldsymbol{C_y} \approx \frac{\tilde{\boldsymbol{Y}}\tilde{\boldsymbol{Y}}^T}{N-1}. \tag{3.10}$$

As it is very important to treat the observations $\boldsymbol{z}(\omega)$ as a random variable ([78]; but also clear from Eq. (2.5.3)), form the observation ensemble matrix $\boldsymbol{Z}$ by sampling the observations RV $\boldsymbol{z}_t(\omega)$. This RV is typically assumed to be Gaussian with covariance matrix $\boldsymbol{C_z}$ and mean $\boldsymbol{z}$

$$\forall i = 1, ..., N : \boldsymbol{z}_i \sim \mathcal{N}(\boldsymbol{z}, \boldsymbol{C_z}). \tag{3.11}$$

It is now possible to write the EnKF update equation as a special case of Eq. (2.49) as

$$\hat{\boldsymbol{X}} = \boldsymbol{X} + \boldsymbol{K}(\boldsymbol{Z} - \boldsymbol{Y}), \tag{3.12}$$

where $\boldsymbol{K}$ is again the Kalman gain defined in Eq. (2.50). It should be pointed out that the resulting updated ensemble directly represents the posterior estimate $\hat{\boldsymbol{x}}(\omega)$. While being clear from the equation, this is one of the nice features of the algorithm: as opposed to, for example, generic particle filter algorithms (*cf.* section 3.4.3.3), no weighting and explicit re-sampling is necessary.

Challenges in Probabilistic LCE Let us first describe the major challenges for probabilistic LCE updating methods. In the next paragraph, these challenges are discussed when reviewing the existing implementation variants. Naturally, like with any Monte Carlo method, a major source of errors in probabilistic LCE updating methods is the limited sizes of the sample. For computational and storage reasons, this has to be kept small[2]. The major consequences of the limited ensemble size are *spurious correlations* in the estimated covariance matrices [170] and the so-called *inbreeding* of the ensemble members [171, 172, 267], both of which may lead to a collapse of the ensemble called *filter divergence* [21, p. 213]. Kovalenko et al. [216] discusses — under some simplifying assumptions — theoretical aspects and the practical influence of sampling error in the EnKF.

[2]In a commercial context, also the amount of simulator licenses available for the solution of the forward problem may be a significant aspect.

Spurious correlations are estimated correlations (*cf.* Eqs. (3.9) and (3.10)) between two distant grid points $\boldsymbol{a}, \boldsymbol{b} \in \mathcal{X}_s$ of the model, which are implausible because of the underlying physical system (*e.g.* correlations between pressure measurements in the atmosphere taken 100s of kilometres apart). An example is given by Houtekamer and Mitchell [171, figure 6]. These *un-physical* correlations can be directly attributed to the Monte Carlo sampling error [170, p. 3279]. Their possible effect was realised already in the work of Julier et al. [194, p. 1629], though, and one of the reasons for the development of the UKF. Evensen [124, p. 101] states that these correlations may also lead to an underestimation of variance.

Inbreeding denotes the use of the same ensemble members in several locations of the update procedure, causing linear coupling between them with successive updates [171, section 2e]. Methods to ameliorate inbreeding include using multiple ensembles with lesser members and exchanging the Kalman gain. The effects are discussed in the works of Houtekamer and Mitchell [170], Houtekamer et al. [173], and Mitchell and Houtekamer [267] and found to be positive for the variance estimation. Another possibility suggested in the publication of Houtekamer and Mitchell [171, p. 808] is to compute N Kalman gains, one for each ensemble member k, and to omit the kth ensemble member to be updated in its computation (*e.g.* [273, 267]). Several approaches of this kind have been reviewed by Mitchell and Houtekamer [267] and applied to the logistic map as a test problem. Each of the advanced approaches was found to outperform simple ones.

Rank deficiency of the estimated covariance is a quite natural issue (as the EnKF may well be seen as a low-rank approximation of the KF; *e.g.* Nerger et al. [275]), which may occur in both the square root scheme and the standard scheme (*cf.* [122, section 7.2], [205]). This can be ameliorated by using a proper pseudo-inversion [288, section 3.6] in the update (*cf.* [122, section 7.3], [6, section 5.1], also discussed in appendix D).

Variants of Probabilistic Linear Conditional Expectation There are two major variants of the probabilistic LCE update: the older *perturbed observations* (PO) EnKF, where the name stresses the fact that the observations $\boldsymbol{z}(\omega)$ have to be treated as RVs (*cf.* [78]) and *square root* (SQ)

implementations which are related to the numerically more stable implementations of the original KF [295, 26, 61]. These two variants are discussed in the next two paragraphs, while lesser known variants and modifications are described in the third paragraph.

In the PO formulation, Eq. (3.12) is almost directly implemented, with some minor modifications to optimise the involved linear algebra operations (*e.g.* [249, section 4.1] or [121, section 4]) and ensure numerical stability (*e.g.* [6, section 5.1], [124, p. 89]). Care must also be taken when implementing this variant for high dimensional observations; here the Sherman-Morrison-Woodbury formula may help (*e.g.* [249, section 4.2], *cf.* [148]). Another possibility is to assimilate uncorrelated observations in batches (*cf.* [172, p. 125, section c]). Anderson [20, section 4c] suggests sorting of ensemble members to limit the magnitude of updates while retaining statistical properties. Hunt et al. [175] show how observations can be assimilated asynchronously to limit the amount of update steps. A discussion on the implementation of an EnKF on parallel computers can be found in the work of Houtekamer and Mitchell [172].

In the square root formulation the mean and covariance are updated independently and 'perturbation' of the observation is not necessary. Therefore these methods are sometimes called *deterministic ensemble Kalman filters*, stretching the fact that the initial ensemble is still a Monte Carlo sample. On the other hand, sampling errors introduced by the perturbed observation variant may be omitted ([268, p. 2807], [170, p. 3273], [14, p. 396]). Instances of the square root implementation variants are the *ensemble transform Kalman filter* (ETKF) [65], the *ensemble adjustment Kalman filter* (EAKF) [19], and the filters developed in [358] and [122]. Tippett et al. [341] put them in a unifying framework and Miyoshi [269] reviews some of these methods. Some variants may introduce a bias (*e.g.* [122]), which is pointed out in the work of Livings et al. [240]. There, a *symmetric square root* implementation is proposed that fixes this problem. A different problem of the square root implementation is that for nonlinear models, the updated ensemble may have a pathological sampling structure (all members except one are clustered in one point, and the single outlier represents the variance; *cf.* [23, p. 4189]), which may be ameliorated with so-called *random rotations*: a multiplication by a mean preserving random orthogonal matrix [124, p. 100]. However, this practically destroys all non-Gaussian structure in the ensemble and therefore must be used with care. A popular SQ variant is the *local* ETKF (LETKF) [285], which scales nicely with the dimension of observations from a com-

putational perspective. Hunt et al. [176, section 4] extend this LETKF towards asynchronous assimilation of observations.

Lesser known formulation include the implementation of Bergemann et al. [51] and Bergemann and Reich [50]. There, a specially crafted *ordinary differential equation* (ODE) is integrated to solve the update equation, allowing for a localisation approach (*cf.* section 3.4.1.4) with square root schemes. Another 'deterministic' implementation is presented by Sakov and Oke [311]. It is more close to the PQ one, and also allows for traditional Schur-product localisation. Heemink et al. [160] propose a square root filter in combination with rank reduction by eigenvalue decomposition. Lermusiaux and Robinson [232] and Lermusiaux [231] introduce a method called *error subspace statistical estimation*, which also employs an ensemble for the integration of the forward model, while the error subspace in the update is represented by a singular value decomposition and thus related to the SEIK/SEEK methods described above [291, 293, 292]. A hierarchical variant called *hierarchical EnKF* (HEnKF), where the estimated covariance and mean are again seen as RVs with certain Bayesian priors, is presented by Myrseth and Omre [272]. It is empirically demonstrated to be more robust for smaller ensemble sizes and designed to have less problems with rank deficiency. Phale and Oliver [290] introduces a *constrained EnKF* to avoid updates which exceed predefined physical limits. However, we see such problems as usually being a result of not using a suitable transformation of the joint model-observation manifold to a vector space (*cf.* section 2.1). Ng et al. [277] discuss the role of model dynamics for the EnKF performance in chaotic systems.

For practical applications of probabilistic LCE updating methods it has been repeatedly shown that complimentary regularisation methods are required, especially concerning the problem of spurious correlations due to the Monte Carlo sampling errors. As already mentioned, the general problem of underestimation of posterior variance by LCE methods is also an issue.

3.4.1.4. Regularisation Methods

Here regularisation methods which aim at ameliorating drawbacks of the LCE methods — and especially the popular probabilistic sampling variants like the EnKF — are reviewed. They either make use of additional

knowledge on the underlying physical system or on the sampling properties of the methods.

Regularised Sampling Methods These approaches are tailored for reducing the unfavourable influences of Monte Carlo sampling in probabilistic LCE updating methods. They do not impose additional knowledge from the problem structure on the equations and simply 'enforce' theoretical relations, thus they are not problem specific regularisation methods.

One of the simplest approaches is to enforce correct mean and variance to ensembles sampled from Gaussian RVs, called *second order correct sampling* and suggested by Pham [292] and Evensen [122, p. 547]. Pham [292, appendix] describe it together with additional linear constraints to eliminate correlations; therefore we describe a simplified (and less exact) approach in appendix D. Pham [292, p. 1205] finds that second order exact sampling '[...] permits the reduction of the number of ensemble members to a strict minimum and yet does not cause any degradation of performance.' Additionally, Zhang et al. [372] shows the similarity of a second-order exact EnKF with independence constraints to square root filters.

Evensen [122, section 4] and Evensen [123, section 11.4.2] propose another 'improved sampling' method. When sampling an ensemble of size N, he suggests to first create a larger ensemble $\beta N, \beta > 1$, and then reduce it by truncated *singular value decomposition* SVD, a proper re-scaling and multiplication with a random orthogonal matrix to the desired size N. The method can be interpreted as the attempt to better span the random space by creating a better conditioned ensemble, thereby explaining more of the variance of the process. Thus it is basically an application of the KLE mentioned earlier, and clearly related to the way of choosing interpolating states in the SEIK method. The main criticism, stated in the context of initial ensemble creation for geological applications, is

> [...] that the property realizations in the initial ensemble will all be relatively smooth and not necessarily consistent with the prior geological model. [14, p. 397]

However, this is obvious from the L_2-optimality of the KLE, and the somehow 'arbitrary' truncation of the expansion in the described algorithm.

Usually, one would decide where to truncate based on the convergence of singular values. Oliver and Chen [283] have an analysis of this approach, as well as developments of related approaches omitting the smoothness problem, which are generally found to be useful. A related approach can be found in the work of Dovera and Della Rossa [107].

Myrseth et al. [273] describe an EnKF variant which resamples the Kalman gain $\boldsymbol{K}$ from a bootstrapped ensemble, obtained from a non-parametric estimation of the cumulative density function of $\boldsymbol{x}$. This approach reduces the inbreeding effect at the price of introducing another source of sampling errors. Zhang and Oliver [370] give a related approach, where bootstrapping is applied to screen the Kalman gain and achieve similar effects as those from localisation (*cf.* section 3.4.1.4). However, this approach is also applicable in case the observation does not have a notion of locality. Zhang and Oliver [371] compare it to classical localisation. Mitchell and Houtekamer [267] describe an approach which uses the jack-knife estimator to obtain both an estimate of the gain and an estimate of its uncertainty, and subsequently produce a different Kalman gain for each ensemble member. This method is found to have superior performance to others, at the cost of increased computational and conceptual effort.

Localisation Often used regularisation methods for LCE implementations are *covariance localisation* and *local updating*. Especially in the context of probabilistic variants like the EnKF, these two approaches are highly popular and a plethora of different methods has been developed. These methods are applicable to systems where it is possible to define a notion of *distance* and *location*. The central idea of both approaches is the same: observations should have only a local area of influence on the model quantities, but due to limited accuracy of estimated covariances spurious correlations over long distances caused by noise may arise Houtekamer and Mitchell [171, p. 808]. Such noisy correlations are removed via localisation by enabling the user to introduce additional knowledge on the observation locality into the updating method. Thus it is obvious that effective localisation approaches are bound to be application specific and must be designed with care (*cf.* [206, p. 1158]).

Covariance localisation uses a tapering function to damp off covariances beyond a certain distance in the estimated covariance matrices

Eqs. (3.9) and (3.10). It is performed by elementwise multiplication (*Schur* or *Hadamard* product) of the covariance matrices by a tapering matrix [172, p. 125]. The tapering functions are usually compactly supported, with the most popular function being Eq. (4.10) of Gaspari and Cohn [134, p. 748], a 5th order piecewise polynomial closely resembling a Gaussian. The covariance localisation approach is straightforward only for the PO variant of EnKF . For square root variants, local updating (see below) is more popular. Petrie [289] developed a covariance localization by Schur product for the ETKF. However, in its current state this implementation does not function as desired. An approach closely related to covariance localisation applies the localisation directly to the Kalman gain and is found to have beneficial effects in some cases [371].

Chapter 5 discusses an approach which combines multi-scale wavelet anaylsis with ideas of covariance localisation.

Local updating solves the updating equation gridpoint by gridpoint[3] and ignores an observation in the update when it is beyond a certain user defined distance (*cf.* [124, p. 101]). This approach is also especially well suited for large observation sets [285]. In its simplest form it must be taken care that observation sets change only gradually while moving along the model space (*cf.* [176, section 2.2.3], [150, p. 2906]). However, methods to gradually decay the influence of observations also exist (*cf.* [176, section 2.3.4]). A local analysis scheme for SEIK is developed in [276], thereby showing that localisation approaches may also be beneficial for non-probabilistic methods. Fukumori [131] develops a local analysis approach for the original Kalman filter.

According to Sakov and Bertino [309, 308] both methods are not equivalent, but practical differences may be insignificant. Localisation approaches are considered as effectively increasing the ensemble size for probabilistic LCE methods [124, p. 84] [172, p. 135] [170, p. 3284]. They also increase the rank of the estimated covariance matrices [39] [281, p. 34]. However, they have been criticised for creating possible dynamical imbalance[4] [268] [241, section 3c] [170, p. 3274] [50]. On the other hand, Oke et al. [281, p. 41] demonstrate with a test problem that a 'large

[3] Updating gridpoint by gridpoint alone makes no difference in the method whatsoever and is equivalent to the 'all-at-once' update (*e.g.* [245, p. 92]).

[4] Dynamical imbalance is a violation of constraints between model variables, caused by updates which do not lie in the linear span of the prior [281, p. 40] [92].

enough' localisation radius only causes insignificant imbalance. Kepert [207] presents a localisation method which aims at reducing this problem for atmospheric and oceanic applications. It is clear that the localisation distance thus becomes a tuning parameter of the algorithm (*cf.* [281, p. 42]), whose optimal value also depends on the ensemble size [87, p. 2]. Adaptive approaches are under development [63, 64]. Emerick and Reynolds [113] discuss a method to estimate a localisation distance from data sensitivity and correlation length of the underlying model. An argument for the validity of localisation is given by Furrer et al. [132, p. 3] via conditional expectations. A combination of covariance localisation and local updating is given by Janjić et al. [183].

Local Averaging A lesser known but related regularisation approach is local averaging [54]. Here the central idea is to filter out local small scale noise from (co-)variance estimates. Therefore it can be interpreted as the complement to the covariance localisation approach — it is essentially a special low-pass filter applied to the estimated covariances. The idea is first thought of for local spatial averaging, but local time averaging is also considered [54, section 7c]. It is related to wavelet-based covariance diagonalisation approaches [286, 99], as they have a similar effect [54, p. 3713]. It also increases the effective sample size [54, p. 3704] and should work hand in hand with covariance localisation approaches [54, p. 3717]. Additionally, optimisation methods for this approach exist [54, p. 3696]. A related idea is the weight interpolation approach for the LETKF presented by Yang et al. [366].

Inflation A widely used regularisation tool, especially in geoscience applications of probabilistic LCE methods, is covariance inflation [24, section 3e]. In its basic implementation, the (cross-) covariance in the estimated matrices Eqs. (3.9) and (3.10) is amplified by a factor $\beta > 1$. Typical empirically chosen values are in the area of 1.01 [124, p. 102]. For probabilistic LCE approaches the optimal value is a function of the ensemble size [151, p. 2789].

Adaptive approaches include the work of Ott et al. [285, p. 421] and hierarchical approaches like the ones given by Anderson [21] and Evensen [124]. Adaptive inflation is probably performed best when the notion of spatial and/or temporal locality is included into the approach, such as in

the works of Anderson [22] and Miyoshi [270].

Mostly, inflation is seen as a way to counteract sampling error effects in probabilistic LCE methods. Hunt et al. [176, p. 120] regard it as some kind of forgetting factor for observations. Houtekamer et al. [173, p. 2127] regard inflation as an opaque method to counteract modelling errors which does not provide additional insight into the structure of this source of errors. However, there is another aspect which should not be forgotten: the fact that *all* LCE methods generally underestimate the 'true' variance of the posterior due to the limitation to being affine functions of the observations could also be counteracted by inflation. This is discussed shortly by Hunt et al. [176, p. 119] in the context of probabilistic methods, but it should be considered an issue for all *linear* CE methods.

Others Johns and Mandel [189] describe a combination of EnKF with *Tikhonov regularisation* to incorporate a priori assumptions about the size and smoothness of the desired solution. This approach is applied by Mandel et al. [251] to a wildland fire model. Bergemann and Reich [52] present an approach which aims at mitigating the imbalance effects of updates. They introduce an approach which basically stretches the assimilation of a measurement over time. Another covariance regularisation approach, complementary to inflation and localisation, is given by Ueno and Tsuchiya [343]. They prescribe knowledge about variable dependence to the *inverse* of the covariance matrix, thereby obtaining a regularised matrix with finite determinant.

3.4.1.5. Diagnosis Methods

Several approaches have been developed for a posteriori diagnosis of LCE estimates. Also here, analysis methods for the probabilistic approaches dominate because of their popularity. Desroziers et al. [101] derive some consistency diagnostics for the covariances of observation, background and estimation errors in observation space. Liu and Kalnay [238] and Li et al. [233] propose a method to diagnose observations with large random errors. Liu et al. [239] present a complimentary method which computes the sensitivity of the posterior to the observations, giving an indication how much information has been 'extracted' from the observations.

The computational cost of such diagnosis methods is almost always practically negligible. Therefore such indicators can be routinely computed and monitored during a workflow.

3.4.2. Approximate Non-linear Methods

Here, methods are discussed which aim to improve over strictly linear methods by means that cannot converge to the full CE solution. Many of the methods are LCE methods which have been extended with iterations, but we also collect here all methods that are based on Gaussian mixtures. While they may converge to a fully non-linear solution when the amount of Gaussian 'basis distributions' goes to infinity, there is no automatic procedure to compute weights, means and covariances [303, p. 32]. Also choosing the amount of Gaussian basis functions is heuristic.

3.4.2.1. Spectral Representation

Naveau et al. [274] describe an extension of the classical Kalman filter to the *closed skew-normal distribution.*An extension to heavy tail distributions is presented by Sornette and Ide [326] with the *Kalman-Lévy* filter. An iterative version of the EKF is given by Jazwinski [187, theorem 8.2].

The work of Alspach and Sorenson [17] contains one of the first descriptions of a *Gaussian sum filter*. There, a weighted sum of Gaussian distributions is used to arbitrarily well approximate other distributions. This can be seen as a non-orthogonal spectral decomposition. An EKF is applied to each of the Gaussians and the output is formed as a convex combination of them. Ito and Xiong [179] apply the approach with a different weight updating rule.

3.4.2.2. Deterministic Sampling Representation

The application of deterministic or quasi-Monte-Carlo integration rules to the solution of spectral update equations can be seen as instances of

this approach. For example, Luo et al. [246] describe the *scaled unscented transform Gaussian sum* filter which combines ideas from Julier and Uhlmann [192] and Alspach and Sorenson [17].

3.4.2.3. Probabilistic Sampling Representation

Sætrom and Omre [336, 337] introduce a non-linear estimation procedure based on EnKF and kernel-shrinkage regression techniques. Basically, they aim at projecting onto a larger set of functions like second order polynomials or second order exponential kernel functions [336, Eq. (11)]. However, by doing so they introduce hyperparameters which have to be chosen. Zupanski [374], Zupanski et al. [375] introduce the *maximum likelihood ensemble filter* (MLEF), a hybrid method based on maximum likelihood and ensemble data assimilation. It creates a maximum likelihood solution and a covariance update, similar to square root algorithms. Gu and Oliver [147] introduce the *ensemble randomised maximum likelihood filter* (EnRML), which is essentially an iterative EnKF (*cf.* [87, p. 3]). Other iterative EnKF are presented by Yang and Kalnay [365] and Sakov et al. [312].

Different approaches to combine Gaussian mixture models with EnKF can be found in the works of Bengtsson et al. [47], Smith [324], Hoteit et al. [169], Stordal et al. [330], and Dovera and Della Rossa [106]. Anderson and Anderson [24] also employ a Gaussian mixture estimated form a Monte Carlo sample. The posterior is computed by explicit convolution of Gaussian kernels (prior Gaussian mixture times Gaussian likelihood), and a new sample is generated.

The *rank histogram filter* of Anderson [23] approximates the prior and likelihood by piecewise constant/linear functions and Gaussian tails. The normalised posterior is formed by explicit multiplication. New ensemble members are placed on an equiprobabilistic grid. Although the computational overhead seems to be quite large, the author reports a successful application to a global numerical weather prediction experiment.

3.4.3. Fully Non-linear Methods

Fully non-linear methods aim at solving the full conditional expectation (or, equivalently, Bayes' formula; see Eq. (2.5.1)). Here, the difficulty is not only the possibility of complex model and measurement operators, but also the fact that the solution is an (almost) *arbitrary* function of the data. However, by doing so these methods (in principle) avoid a systematic underestimation of the posterior variance. Such methods are also called 'non-parametric' (*cf.* [265]), since they do not try to fit parameters of assumed distributions but allow for arbitrary ones.

A very good review of non-linear filters is in the article by Daum [96]. Another review of non-linear methods is given by Budhiraja et al. [77].

3.4.3.1. Spectral Representation

Typical methods involve the application of classical numerical methods to the explicit solution of certain partial differential equations (PDEs) describing the evolution of incomplete information, and its conjunction with observations, over time. Examples are methods for solving the *Stratonovich-Kushner* equation [331] for the recursive evolution of the normalised conditional probability density, such as presented by Lototsky et al. [243], Lototsky [244] and Bain and Crisan [36, section 8.4]. A related approach for solving the *Zakai* equation [367] for the non-normalised conditional probability density is described by Bain and Crisan [36, section 8.5]. Also used are numerical solutions to the *Fokker-Planck* [302] or *Kolmogorov-forward* equation [264, 200].

Strictly speaking, these are approaches tied to the representations of RVs which have been not considered further (*cf.* section 3.2).

A quite different approach is taken by El Moselhy and Marzouk [110]. The authors aim to arrive at an approximation to the Bayesian posterior by constructing a special map which creates the posterior as the push-forward measure of the prior (*cf.* appendix A.3) using standard orthogonal polynomials. The map is computed by solving an optimisation problem.

3.4.3.2. Deterministic Sampling Representation

These methods are sometimes subsumed under the name *deterministic particle filters*. However, usually the name particle filter is reserved for the probabilistic ones (see section 3.4.3.3 below).

Examples of such methods are the *point mass filter* introduced by Bergman et al. [53]. An peplication to GPS navigation is given by Lehmann [228]. Another approach is presented by Kalender and Schöttl [197], where sparse grids in six dimensions are used for the real-time integration of the Fokker-Planck equation.

3.4.3.3. Probabilistic Sampling Representation

Since the seminal paper of Gordon et al. [143], the arguably most successful class of methods to solve the non-linear filtering problem are *particle filters* (PF). They are sequential Monte Carlo methods [261, 229] specifically adopted to approximate solutions of the full recursive conditional expectation (*cf.* section 2.5.3). Again, a plethora of different variants has been proposed (*cf.* [96, Table II]; [143, 347, 346, 215, 349, 104]). A highly efficient particle filter which 'nudges' the particles towards observations is given by van Leeuwen [348]. Another highly efficient approach is described by Daum and Huang [97], where a PDE is used to let particles 'flow [...] to the correct region of state space for the computation of Bayes' rule'.

Recent books on the subject of particle filters include the works of Doucet et al. [105] and Ristic et al. [303]. A tutorial for online applications is given by Arulampalam et al. [32]. Bashi et al. [41] present notes on distributed implementations. Chen [90] wrote a review article on the larger topic of Bayesian filtering. A review for geophysical applications is given by van Leeuwen [350].

Particle filters are currently the arguably only methods applicable to 'high-dimensional', dynamic, fully non-linear CE problems. However, it needs to be pointed out that, contrary to common belief, also particle filters suffer from the 'curse of dimensionality'. They have to be designed, implemented and tuned carefully (as pointed out by Daum [96, p. 61] and

Ristic et al. [303, section 3.6]), and the choice of the proposal density is crucial for their performance (with [97] being a possible exception).

3.4.4. Related and Derived Methods

Here related popular methods are shortly presented which, for various reasons, could not be integrated in above review structure.

3.4.4.1. Markov Chain Monte Carlo Methods

The class of Markov Chain Monte Carlo (MCMC) methods — including the celebrated *Metropolis-Hastings* (MH) method [261, 262, 158, 237, 91] and the *Gibbs* method [136, 351] — are also popular for solving inverse problems [340, chapter 2]. They may be combined with surrogate models or reduced spectral representations for the solution of the forward problem [255, 28, 253, 254, 219, 139]. However, they are usually not suited for sequential estimation: in MCMC methods, observations are typically included in one go over the whole evolution time of the system of interest. One reason for this is that the prior is usually easy to sample from, whereas the posterior is not [340, p. 52]. However, in a sequential setting the posterior of one step — integrated forward in time — is the prior of the next. Therefore these methods are — in contrast to particle filters — not recursive. This may not always be feasible and, with closed loop identification and control in mind, also not desirable.

Additional drawbacks of MCMC methods include the fact that samples are not independent, which makes estimating any statistics (except the mean) from them problematic. Approaches towards mitigating this problem include [43]. The independence property of samples is also crucial for usual statistical formulas for the spread or accuracy of the estimates. Finally, MCMC methods need a burn-in period during which all samples have to be discarded, as the equilibrium distribution is only the asymptotic distribution.

3.4.4.2. Smoothers

A class of methods which is closely related to the filters reviewed in the previous section are the *smoothers* (*e.g.* [209, 210]). These methods basically propagate information obtained from observations backwards in time by extending the model state vector at time t with the vectors from time $t-1, t-2, ..., 0$. An example is the *ensemble Kalman smoother* (EnKS) [125],[123, section 9.6], which is related to the EnKF. One example is given by Kalnay et al. [199] for the LETKF scheme (then called 4D-LETKF).

Smoothers can basically be constructed from every filter [265] (or, as some see it, the other way around). Due to this direct relationship, smoothers are not considered separately.

3.4.4.3. Variational Data Assimilation

Variational data assimilation schemes like 3DVAR and 4DVAR [48] can be seen as a least squares approach to the incorporation of observations into dynamic model variables [335, p. 474f]. They are highly popular with the athmospheric and oceanic sciences (*e.g.* [297]), where they have been developed. However, they are not suited very well to joint parameter-state estimation problems ([124, p. 85] and references [26,27,8] therein). Additionally, LCE methods as well as hybrids with variational schemes [150] are entering the application area of atmospheric and oceanic sciences and are quite competitive in operational settings (*e.g.* [173]). See the work of Lahoz et al. [220] for a recent collection on the topic.

3.4.4.4. Regularised Optimisation

Another large group of methods which is closely related to variational schemes is *regularised optimisation* [115]. These methods treat the inversion problem as an ill-posed optimisation problem that has to be regularised with additional information. Popular methods are *Tikhonov regularisation* and *Landweber iterations*. Their result is a single, in some sense optimal estimator.

Regularised optimisation problems can be seen as a special case of the

more general problem we are considering in the incomplete information setting (*cf.* [335]). For complex applications, they are also used (*e.g.* [320, 34]), but we think that the incomplete information approach is better suited as it provides much more information. However, regularisation methods are highly popular in certain areas, for example image reconstruction — where obtaining a single best estimate is somewhat compulsory.

3.5. Discussion

In this chapter methods for the solution of stochastic inverse problems have been reviewed. It has become clear that linear conditional expectation methods are highly promising — though their variance estimates should be treated with due diligence and care. Especially probabilistic LCE methods like the EnKF are quite well developed. However, they have to be combined with application specific regularisation approaches to be able to show their full power.

Since most of the problems of probabilistic LCE methods stem from sampling errors it is of practical interest to develop non-probabilistic methods. Additionally, such methods should be able to maintain a constant subspace span (or even detect 'optimal' subspaces), and not possibly result in pathological ensembles during the updating process like, for example, ensemble square root filters — all of this while maintaining the applicability to high dimensional problems. A methodological development into this direction is discussed in the next chapter, where a sampling-free spectral LCE approach is presented.

Chapter 4.

Linear Bayesian Inference Using a Spectral Series Expansion

Being one example of a spectral series expansion suitable for the representation of random variables, it is demonstrated how the polynomial chaos expansion can be used to derive a sampling-free, computationally efficient linear Bayesian update formula. The central idea is to combine the (recursive) LCE Eq. (2.49) with the PCE representation Eq. (3.2). In the following this approach will be denoted as *Linear Polynomial Chaos Updating* (LPCU). Related publications can be found in [8, 7, 6, 10, 11]. The specific, new implementation introduced in the following is the *Square Root Polynomial Chaos Updating* (SRPCU).

It is worth pointing out that this approach is not limited to PCE representations — since Eq. (2.49) is expressed in RVs, all kinds of representations are usable, thereby leading to a large class of methods (*cf.* [245, chapter 4]). A well-known example are the probabilistic implementations of the ensemble Kalman filter family (*cf.* section 3.4.1.3), and also other stochastic series expansions should be readily usable (*cf.* section 3.2).

4.1. Motivation

From the literature review in the previous chapter it is evident that the vast majority of currently used LCE implementations is based on a stochastic sampling representation of the involved RVs. The inevitable sampling errors fostered the development of square root implementations,

which partly avoid this problem but have other specific issues [354, 124]. Additionally, regularisation approaches have been developed which mainly target sampling errors. Unfortunately, none of these methods succeeds at completely avoiding them. Naturally, this is different for spectral representations of RVs: there, no sampling is necessary at any stage — instead, they introduce a truncation error. A second drawback of the stochastic sampling representation is its slow convergence. For spectral approaches like PCE this is different; while not being without issues (*e.g.* [126]) they possess up to exponential convergence [364], meaning that a larger share of variance may be represented with fewer coefficients — clearly a favourable feature for large scale applications.

An advantage of spectral representation implementations of the LCE update is that it — naturally — directly integrates with recently popular spectral uncertainty quantification methods and avoids the detour to probabilistic sampling. Another advantage is directly connected to applications with dynamical systems: since the posterior is again represented by a spectral expansion, sequential updating schemes are conceptually very easy — in contrast to, for example, MCMC methods.

4.2. Related Work

Related approaches have been independently developed by Saad [307, p. 52–55], Blanchard et al. [66] and Zeng and Zhang [369]. The former two are derived as extensions of extended Kalman filter theory, whereas the latter one is presented as some kind of modified EnKF. However, they basically have the same underlying idea.

The major difference in the formulation given here is that it avoids some of the more problematic issues inherent in these already existing approaches by employing a square root form of the estimator.

4.3. Linear Polynomial Chaos Updating

With the presentation of the LCE in section 2.5.4 and the PCE in section 3.2.1.1 the description of the approach is almost trivial.

Due to the mutual orthogonality of the PC basis functions one may directly project Eq. (2.49) onto the PCE using Eq. (3.3). Remember that the multi-index set $\mathcal{J}$ is used as the representation for the multivariate Hermite basis (see Eq. (3.2)). Thus, projection of Eq. (2.49) leads to

$$\forall \alpha \in \mathcal{J} \,:\, \hat{\boldsymbol{x}}^{\alpha} = \boldsymbol{x}^{\alpha} + \boldsymbol{K}(\boldsymbol{z}^{\alpha} - \boldsymbol{y}^{\alpha}), \tag{4.1}$$

an updating equation for the vector of coefficients for each of the basis functions. In terms of the Hermite transform (see appendix B.4), Eq. (4.1) can be compactly expressed as

$$\mathscr{H}(\hat{\boldsymbol{x}}) = \mathscr{H}(\boldsymbol{x}) + \boldsymbol{K}(\mathscr{H}(\boldsymbol{z}) - \mathscr{H}(\boldsymbol{y})). \tag{4.2}$$

Note that for simplicity the modelling error of the measurement operator, $\eta_t(\omega)$, has been assumed as zero — in many practical applications this is done anyway. However, it may be readily re-introduced if necessary.

4.3.1. Numerical Implementation

After a suitable selection of a finite basis index set $\mathcal{J}_Z$ (truncation, *cf.* section 3.2.1.2), a simple numerical representation of the — now finite — expansion is to collect all the column vectors of coefficients into matrices, *i.e.*

$$\boldsymbol{X} = [..., \boldsymbol{x}^{\alpha}, ...],\ \boldsymbol{Y} = [..., \boldsymbol{y}^{\alpha}, ...],\ \boldsymbol{Z} = [..., \boldsymbol{z}^{\alpha}, ...]. \tag{4.3}$$

Note that due to the joint model-data space, the spectral basis is the same for the model and the data space. With these matrices the update Eq. (4.1) simply reads as

$$\hat{\boldsymbol{X}} = \boldsymbol{X} + \boldsymbol{K}(\boldsymbol{Z} - \boldsymbol{Y}), \tag{4.4}$$

a direct linear update equation for the polynomial chaos coefficients. The first remaining ingredient is the computation of the Kalman gain matrix $\boldsymbol{K}$:

$$\boldsymbol{K} = \boldsymbol{C}_{\boldsymbol{xy}} \left(\boldsymbol{C}_{\boldsymbol{y}} + \boldsymbol{C}_{\boldsymbol{z}}\right)^{-1}. \tag{4.5}$$

Computing the Kalman gain involves the estimation of two involved (cross-) covariance matrices $\boldsymbol{C_{xy}}$ and $\boldsymbol{C_y}$ from the truncated PCE representations of the RVs. This is straightforward using the Hermite integration algebra (*cf.* appendix B.3; the explicit formula for a complete cross-covariance matrix is given in Eq. (B.22)). However, introducing the diagonal Gram matrix

$$(\boldsymbol{\Delta})_{\alpha\beta} = \mathbb{E}\left(H_\alpha H_\beta\right) = \mathrm{diag}(\alpha!), \tag{4.6}$$

the (cross-) covariance between two RVs represented by PCE coefficient matrices can be compactly expressed. For this, the following abbreviation[1] is helpful:

$$\tilde{\boldsymbol{X}} = \boldsymbol{X} \setminus \bar{\boldsymbol{X}}, \tag{4.7}$$

and similar for $\boldsymbol{Y}$. These are matrices with the first column of coefficients — corresponding to the constant Hermite polynomial and therefore to the expectation of the RV — being omitted. With this, the cross-covariance between model and data variables can be conveniently estimated by

$$\boldsymbol{C_{xy}} = \tilde{\boldsymbol{X}}\boldsymbol{\Delta}\tilde{\boldsymbol{Y}}^T, \tag{4.8}$$

and similar for $\boldsymbol{C_y}$.

Additionally, computing the Kalman gain involves the inversion of $\boldsymbol{C_z} + \boldsymbol{C_y}$. However, depending on the involved PCE truncation, $\boldsymbol{C_y}$ may not have full rank. If the measurement error represented by $\boldsymbol{C_z}$ has a small variance, $\boldsymbol{C_z} + \boldsymbol{C_y}$ may become close to singular and inversion[2] may be unstable or impossible. Therefore the inverse in Eq. (4.5) is replaced by the *Moore-Penrose pseudo inverse* (*cf.* [322, 142]), which is numerically favourable and produces a least-squares approximation in case of singularity. Typically this pseudo-inverse is computed from a singular value decomposition, though other methods exist [322]. The computational cost of the SVD is usually acceptable.

The last remaining ingredient is related to computing the difference $\boldsymbol{Z}-\boldsymbol{Y}$. Unfortunately this turns out to be a critical one. In section 2.5.3 it is assumed that the individual measurement errors $\boldsymbol{\varepsilon}_t(\omega)$ are uncorrelated in time (to be able to sequentially process the observations). Additionally,

[1] This corresponds to Eq. (A.16) in the appendix.

[2] Clearly one would solve a system of equations rather than performing the inversion, but this results in the same problems.

in section 2.5.5 it is assumed that $\boldsymbol{x}(\omega)$ (and therefore $\boldsymbol{y}(\omega)$) and $\boldsymbol{z}(\omega)$ are uncorrelated. This implies that the primary Gaussian RVs θ_i of the PCE expansions of $\boldsymbol{x}(\omega)$ and $\boldsymbol{z}(\omega)$ must be uncorrelated — and therefore independent. The only consistent way to achieve this is to extend the PCE basis with the necessary number of θ_is for each observation which is to be included in an LPC update. This causes the PCE basis to grow significantly over time (in the context of sequential updates) and quicklys result in computational problems. However, let us denote this approach as the 'correct' one for later reference. There are several possibilities to remedy or circumvent this growing basis: using an adaptive choice of basis may re-compact the PCE expansion after an update when necessary. Another approach is to simply re-use the same θ_is, thereby accepting the additional error: since the measurement error and the model variables are now 'colinear', the estimate which results from the LPCU will have too large variance[3]. This is the approach used by Saad [307]. Zeng and Zhang [369] and Blanchard et al. [66] ignore the additional variability in the calculation of the difference and set all PCE coefficients of $\boldsymbol{Z}$ — except the ones for the mean — to zero, arguing that the '[...] measurement errors are independent of the system state [...]' [369]. This 'independent' observation approach results in a too low variability, as has been shown in the case of the ensemble Kalman filter by Burgers et al. [78]. Due to the significant computational advantages and given that adaptive choices of PCE basis are still a research topic, the two last approaches will also be considered further — although they are, strictly speaking, incorrect. This is demonstrated by numerical examples in section 4.5. The thee implementation variants ('correct', 'colinear', and 'independent' treatment of observations) of the LPCU approach are summarised in algorithm 1.

However, an additional approach is introduced in the following that is based on an alternative formulation of the update equation. It circumvents the extension of the PCE basis in a consistent way.

[3]Intuitively, this can be explained as follows: subtracting two vectors pointing into the same direction results in a shorter vector than subtracting two orthogonal vectors of the same length — this means too small innovations result from $\boldsymbol{Z} - \boldsymbol{Y}$, and therefore a too small LPC update.

Algorithm 1 The linear polynomial chaos update (LPCU)

procedure LPCU($\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{C_z}$)
  $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) \leftarrow$ ExtendBasis($\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$) ▷ Optional: 'correct' obs.
  $\tilde{\boldsymbol{X}} \leftarrow \boldsymbol{X} \setminus \bar{\boldsymbol{X}}$ ▷ Remove the $\alpha = 0$ coefficients
  $\tilde{\boldsymbol{Y}} \leftarrow \boldsymbol{Y} \setminus \bar{\boldsymbol{Y}}$
  $\boldsymbol{C_{xy}} \leftarrow \tilde{\boldsymbol{X}} \boldsymbol{\Delta} \tilde{\boldsymbol{Y}}^T$
  $\boldsymbol{C_y} \leftarrow \tilde{\boldsymbol{Y}} \boldsymbol{\Delta} \tilde{\boldsymbol{Y}}^T$
  $\boldsymbol{P}^\dagger \leftarrow$ PseudoInv($\boldsymbol{C_y} + \boldsymbol{C_z}$)
  $\boldsymbol{K} \leftarrow \boldsymbol{C_{xy}} \boldsymbol{P}^\dagger$
  $\boldsymbol{Z} \leftarrow \boldsymbol{Z} \cdot \delta_{0i}$ ▷ Optional: 'independent' obs.
  $\hat{\boldsymbol{X}} \leftarrow \boldsymbol{X} + \boldsymbol{K}(\boldsymbol{Z} - \boldsymbol{Y})$ ▷ This is problematic, as discussed
  return $\hat{\boldsymbol{X}}$
end procedure

4.4. A Square Root Formulation

Square root formulations have been popular mainly due to their advantages in numerical stability [61]. They have been first conceived for the original Kalman filter [295, 26]. However, square root approaches have also become popular for probabilistic LCE (*cf.* section 3.4.1.3), mainly because they do not include additional sampling errors into the process via an observation ensemble. It is worth pointing out at the beginning that generally, square root implementations and their 'standard' counterparts are *non-equivalent* — with equivalence only for purely multivariate Gaussian RVs. The non-equivalence is also demonstrated later by a numerical example.

Square root approaches compute their update in two steps: the constant and the varying part of the RV are updated independently. The $\alpha = 0$ term in Eq. (4.1) is the Kalman update for the mean [245]. The same conclusion is reached by taking the expectation of Eq. (2.49). This update for the mean remains for the square root formulation:

$$\hat{\boldsymbol{x}}^0 = \boldsymbol{x}^0 + \boldsymbol{K}(\boldsymbol{z}^0 - \boldsymbol{y}^0). \tag{4.9}$$

While the mean is updated using the usual Kalman gain, the major difference lies in the computation of the update for the varying part. A possible formulation for this is described in the following.

Remember from the previous section that the covariance of an RV represented by PCE coefficients may be conveniently written as

$$\boldsymbol{C_x} = \tilde{\boldsymbol{X}} \boldsymbol{\Delta} \tilde{\boldsymbol{X}}^T. \tag{4.10}$$

Due to the diagonal structure of the Gram matrix, it is easy to define $\sqrt{\boldsymbol{\Delta}} := \mathrm{diag}(\sqrt{\alpha!})$. Using a short hand $\boldsymbol{S} = \tilde{\boldsymbol{X}}\sqrt{\boldsymbol{\Delta}}$ it is clear that

$$\boldsymbol{C_x} = \boldsymbol{S}\boldsymbol{S}^T, \tag{4.11}$$

thereby leading to a very specific decomposition of the covariance matrix called a *matrix square root*, $\boldsymbol{S}$. This observation is the key to the square root formulation. We now make the following ansatz: if one could find a matrix $\boldsymbol{A}$ (here the orthogonal matrix $\boldsymbol{T}$ is given later) with

$$\hat{\boldsymbol{S}} = \boldsymbol{S}\boldsymbol{A}\boldsymbol{T}, \tag{4.12}$$

one would have arrived at a linear transformation of the prior covariance square root into the posterior covariance square root — and hence at a representation of the posterior covariance matrix — since

$$\boldsymbol{C}_{\hat{\boldsymbol{x}}} = \hat{\boldsymbol{S}}\hat{\boldsymbol{S}}^T. \tag{4.13}$$

From the matrix square root point of view, $\boldsymbol{T}$ may theoretically be any orthonormal matrix of suitable dimension, since multiplication by such a matrix will not change the covariance matrix $\boldsymbol{C}_{\hat{\boldsymbol{x}}}$. Also note that in Eq. (4.12) the *pre*-multiplication of $\boldsymbol{S}$ with a different transformation matrix, as well as the general linear transformation form of $\hat{\boldsymbol{S}} = \boldsymbol{A}_1\boldsymbol{S}\boldsymbol{A}_2\boldsymbol{T}$, are possible and represent related ways of derivation which are not pursued further here.

Let us proceed by remembering the relation for the updated covariance matrix (*e.g.* [245, p. 96]):

$$\begin{aligned}\boldsymbol{C}_{\hat{\boldsymbol{x}}} &= \boldsymbol{C_x} - \boldsymbol{K}\boldsymbol{H}\boldsymbol{C_x} \\ &= \boldsymbol{C_x} - \boldsymbol{C_x}\boldsymbol{H}^T\left(\boldsymbol{H}\boldsymbol{C_x}\boldsymbol{H}^T + \boldsymbol{C_z}\right)^{-1}\boldsymbol{H}\boldsymbol{C_x}.\end{aligned}$$

Substituting Eq. (4.11) leads to

$$\begin{aligned}&= \boldsymbol{S}\boldsymbol{S}^T - \boldsymbol{S}\boldsymbol{S}^T\boldsymbol{H}^T\left(\boldsymbol{H}\boldsymbol{S}\boldsymbol{S}^T\boldsymbol{H}^T + \boldsymbol{C_z}\right)^{-1}\boldsymbol{H}\boldsymbol{S}\boldsymbol{S}^T \\ &= \boldsymbol{S}\boldsymbol{M}\boldsymbol{S}^T\end{aligned}$$

with $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{S}^T\boldsymbol{H}^T\left(\boldsymbol{H}\boldsymbol{S}\boldsymbol{S}^T\boldsymbol{H}^T + \boldsymbol{C}_{\boldsymbol{z}}\right)^{-1}\boldsymbol{H}\boldsymbol{S}$. A matrix square root of $\boldsymbol{M}$ such that $\boldsymbol{M} = \boldsymbol{A}\boldsymbol{A}^T$ would be a solution of the transform Eq. (4.12), leading to the desired scheme.

Adopting a way of derivation related to the approach of Evensen [122, 124] for a square root formulation of the EnKF, start with an eigenvalue decomposition of $\left(\boldsymbol{H}\boldsymbol{S}\boldsymbol{S}^T\boldsymbol{H}^T + \boldsymbol{C}_{\boldsymbol{z}}\right) = \boldsymbol{B}\boldsymbol{\Lambda}\boldsymbol{B}^T$, which leads to

$$\begin{aligned}\boldsymbol{M} &= \boldsymbol{I} - \boldsymbol{S}^T\boldsymbol{H}^T\left(\boldsymbol{B}\boldsymbol{\Lambda}\boldsymbol{B}^T\right)^{-1}\boldsymbol{H}\boldsymbol{S}\\ &= \boldsymbol{I} - \boldsymbol{S}^T\boldsymbol{H}^T\boldsymbol{B}\boldsymbol{\Lambda}^{-1}\boldsymbol{B}^T\boldsymbol{H}\boldsymbol{S}\\ &= \boldsymbol{I} - (\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{B}^T\boldsymbol{H}\boldsymbol{S})^T(\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{B}^T\boldsymbol{H}\boldsymbol{S})\\ &= \boldsymbol{I} - \boldsymbol{W}^T\boldsymbol{W}\end{aligned}$$

with $\boldsymbol{W} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{B}^T\boldsymbol{H}\boldsymbol{S}$. $\boldsymbol{B}^T$ rotates the simulated measurements into directions aligned with the covariance structure $\boldsymbol{H}\boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{H}^T + \boldsymbol{C}_{\boldsymbol{z}}$, while $\boldsymbol{\Lambda}^{-\frac{1}{2}}$ weights them accordingly. It is exactly here where the additional information (in the form of $\boldsymbol{C}_{\boldsymbol{z}}$) enters the update: $\boldsymbol{C}_{\boldsymbol{z}}$ specifies directions and magnitude of uncertainty (variance) reduction induced by the observations. This information needs to be 'transported' from the data space to the model space (*cf.* section 2.1). However, these two spaces clearly are not the same (the measurement operator stands 'in between') and may not be of the same dimension — but since they are part of the same *joint* space, they are described by coefficients for the same orthogonal spectral basis functions. Thus it is possible to identify the directions and magnitude of variance reduction in data space and apply this variance reduction in model space. This can be accomplished by finding an orthonormal basis for the columns of $\boldsymbol{W}$, which is given by the singular value decomposition $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T = \boldsymbol{W}$. Note that it is necessary to perform a full SVD, such that $\boldsymbol{V} \in \mathbb{R}^{p-1\times p-1}$ with p being the number of PCE terms (including the mean). The SVD leads to

$$\begin{aligned}\boldsymbol{M} &= \boldsymbol{I} - (\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)^T\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\\ &= \boldsymbol{I} - \boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\boldsymbol{V}^T\\ &= \boldsymbol{V}\left(\boldsymbol{I} - \boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\right)\boldsymbol{V}^T.\end{aligned}$$

Since $\boldsymbol{I} - \boldsymbol{\Sigma}^T\boldsymbol{\Sigma}$ is a diagonal matrix — and as $\boldsymbol{C}_{\hat{\boldsymbol{x}}} \geq 0$, so is $\boldsymbol{M}$, hence $0 \leq \boldsymbol{\Sigma}^T\boldsymbol{\Sigma} \leq \boldsymbol{I}$ — it is now easy to compute the desired square root of $\boldsymbol{M}$

$$\boldsymbol{M} = \boldsymbol{V}\sqrt{\boldsymbol{I} - \boldsymbol{\Sigma}^T\boldsymbol{\Sigma}}\left(\boldsymbol{V}\sqrt{\boldsymbol{I} - \boldsymbol{\Sigma}^T\boldsymbol{\Sigma}}\right)^T$$

and the derivation is almost finished with setting

$$\boldsymbol{A} = \boldsymbol{V}\sqrt{\boldsymbol{I} - \boldsymbol{\Sigma}^T\boldsymbol{\Sigma}}.$$

With this result the ansatz Eq. (4.12) becomes

$$\hat{\boldsymbol{S}} = \boldsymbol{S}\boldsymbol{V}\sqrt{\boldsymbol{I} - \boldsymbol{\Sigma}^T\boldsymbol{\Sigma}}\boldsymbol{T}. \tag{4.14}$$

It remains to choose $\boldsymbol{T}$. In *ensemble* square root schemes, a symmetric redistribution using the orthonormal matrix $\boldsymbol{V}^T$ is necessary to arrive at an unbiased scheme [310, 240]. Unbiasedness is not an issue for LPCU, since the mean cannot be changed by this second part of the update, but also here the choice of $\boldsymbol{T}$ is significant: $\boldsymbol{V}$ represents a mapping between the normalised PCE coefficient space and a specific space aligned with the covariance structure given by $\boldsymbol{H}\boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{H}^T + \boldsymbol{C}_{\boldsymbol{z}}$. In that space, the uncertainty reduction according to $\sqrt{\boldsymbol{I} - \boldsymbol{\Sigma}^T\boldsymbol{\Sigma}}$ is performed. Thus, the natural choice $\boldsymbol{T} = \boldsymbol{V}^T = \boldsymbol{V}^{-1}$ maps the result back into the original PCE coefficient space as required. This turns Eq. (4.14) into a linear, symmetric 'covariance contraction' operation. In other words:

- $\boldsymbol{V}$ rotates the normalised prior PCE coefficients $\boldsymbol{S}$ into directions aligned with the additional information represented by the singular values computed above. Note again that this requires the prior and the measurement PCE to be formulated in the same basis. Note that the directions are *sorted by importance* (as usual with the SVD), where importance is defined as 'amount of variance reduction induced by the data'.

- $\sqrt{\boldsymbol{I} - \boldsymbol{\Sigma}^T\boldsymbol{\Sigma}}$ is the diagonal matrix which applies the variance reduction in these directions. The reduction is performed according to the singular values computed from the measurement PCE,

- and $\boldsymbol{V}^{-1} = \boldsymbol{V}^T$ maps the re-scaled directions back into the original basis, resulting in the normalised posterior PCE.

Therefore, Eq. (4.14) turns into

$$\hat{\boldsymbol{S}} = \boldsymbol{S}\boldsymbol{V}\sqrt{\boldsymbol{I} - \boldsymbol{\Sigma}^T\boldsymbol{\Sigma}}\boldsymbol{V}^T. \tag{4.15}$$

To obtain the unnormalised posterior PCE coefficients, set

$$\hat{\hat{\boldsymbol{X}}} = \frac{\hat{\boldsymbol{S}}}{\sqrt{\boldsymbol{\Delta}}}, \tag{4.16}$$

and finally complete the posterior PCE by re-introducing the $\alpha = 0$ term $\hat{\bar{\boldsymbol{X}}}$ computed previously according to Eq. (4.9):

$$\hat{\boldsymbol{X}} = \left[\hat{\bar{\boldsymbol{X}}}\,\hat{\tilde{\boldsymbol{X}}}\right]. \tag{4.17}$$

The square root formulation of the LPCU approach is summarised in algorithm 2.

Algorithm 2 The square root polynomial chaos update (SRPCU)

procedure SRPCU($\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{C_z}$)
  $\tilde{\boldsymbol{X}} \leftarrow \boldsymbol{X} \setminus \bar{\boldsymbol{X}}$ ▷ Remove the $\alpha = 0$ coefficients
  $\tilde{\boldsymbol{Y}} \leftarrow \boldsymbol{Y} \setminus \bar{\boldsymbol{Y}}$
  $\boldsymbol{C_{xy}} \leftarrow \tilde{\boldsymbol{X}}\boldsymbol{\Delta}\tilde{\boldsymbol{Y}}^T$
  $\boldsymbol{C_y} \leftarrow \tilde{\boldsymbol{Y}}\boldsymbol{\Delta}\tilde{\boldsymbol{Y}}^T$
  $\boldsymbol{P} \leftarrow \boldsymbol{C_y} + \boldsymbol{C_z}$
  $\boldsymbol{P}^\dagger \leftarrow \text{PseudoInv}(\boldsymbol{P})$
  $\boldsymbol{K} \leftarrow \boldsymbol{C_{xy}}\boldsymbol{P}^\dagger$
  $\hat{\bar{\boldsymbol{X}}} \leftarrow \bar{\boldsymbol{X}} + \boldsymbol{K}(\bar{\boldsymbol{Z}} - \bar{\boldsymbol{Y}})$ ▷ Update the mean
  $\boldsymbol{S} \leftarrow \tilde{\boldsymbol{X}}\sqrt{\boldsymbol{\Delta}}$ ▷ Normalise the PCE coefficients $\alpha \neq 0$
  $\boldsymbol{B}\boldsymbol{\Lambda}\boldsymbol{B}^T \leftarrow \text{EigenDecomp}(\boldsymbol{P})$
  $\boldsymbol{W} \leftarrow \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{B}^T\tilde{\boldsymbol{Y}}$
  $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V} \leftarrow \text{SingularValueDecomp}(\boldsymbol{W})$
  $\hat{\boldsymbol{S}} \leftarrow \boldsymbol{S}\boldsymbol{V}\sqrt{\boldsymbol{I} - \boldsymbol{\Sigma}\boldsymbol{\Sigma}^T}\boldsymbol{V}^T$
  $\hat{\tilde{\boldsymbol{X}}} = \boldsymbol{S}/\sqrt{\boldsymbol{\Delta}}$ ▷ Remove normalisation
  $\hat{\boldsymbol{X}} = \left[\hat{\bar{\boldsymbol{X}}}\,\hat{\tilde{\boldsymbol{X}}}\right]$ ▷ Concatenate results
  return $\hat{\boldsymbol{X}}$
end procedure

The spectral square root formulation has the distinct advantage that it avoids to extend the spectral expansion by additional RVs when updating. The posterior expansion is obtained by direct incorporation of the additional information into the prior expansion. In this context it is even unnecessary to form a spectral expansion for the observation noise: only the observation, $\boldsymbol{z}$, and the associated covariance structure induced by the error model, $\boldsymbol{C_z}$, are used as input to the update. From a different point of view, this is also the major drawback of the presented approach: any specific structure of the distribution of the error model $\boldsymbol{\varepsilon}(\omega)$ is not used in the update, as opposed to the non-square-root approaches.

Note that in above derivation, the measurement operator $h = \boldsymbol{H}$ has been assumed to be *linear*. However, in the same sense that in ensemble methods the ensemble 'linearises' a non-linear operator (*cf.* [172, Eqs. (2),(3)], [339]), one may say that the PCE representation linearises it by replacing

$$\boldsymbol{C_x H}^T = \boldsymbol{C_{xy}}, \tag{4.18}$$

and similar for all other occurrences of $\boldsymbol{H}$. Therefore, the presented method should also be applicable when the measurement operator is non-linear, as long as

$$h(\mathbb{E}\left(\boldsymbol{x}(\omega)\right)) \approx \mathbb{E}\left(h(\boldsymbol{x}(\omega))\right) \tag{4.19}$$

holds approximately and the variance of $\boldsymbol{x}(\omega)$ is not too large compared to[4] the non-linearity of $h(\cdot)$.

4.5. Numerical Examples

The performance of the SRPCU approach is evaluated on several Bayesian inference tasks of increasing complexity:

- demonstration of the difference between several variants of linear Bayesian updating and full Bayesian updating,
- state estimation for a simple, scalar random walker,
- state estimation and reliability analysis for the non-linear, chaotic Lorenz-63 model, and
- combined parameter-state estimation for the Lorenz-63 model with linear and non-linear measurement operator.

On each problem SRPCU is compared to related, established methods like LPCU, KF, EnKF, EnSRF, and Makov Chain Monte Carlo (MCMC). All

[4] Obviously, this is to be understood as a heuristic argument. Unfortunately, a rigorous consideration is beyond the scope of the present work, but numerical examples in this direction are provided below.

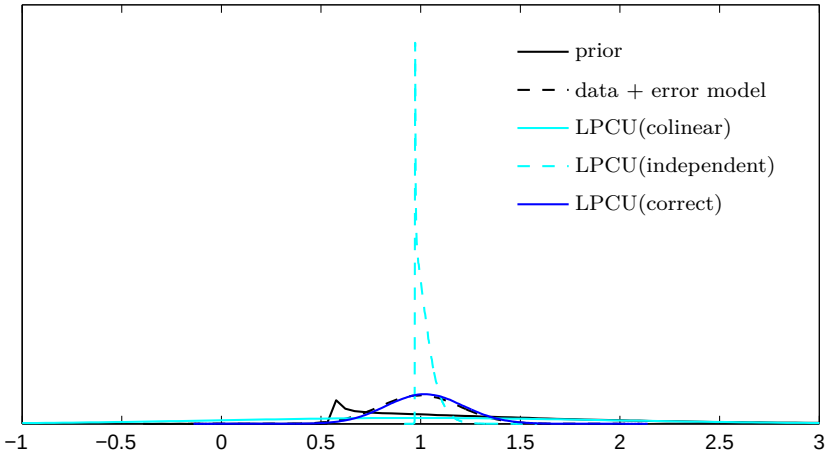

Figure 4.1.: Comparison of different LPCU variants on a scalar update. Shown are estimates obtained by a kernel density technique from 10^6 samples of the PCE.

details of the models, their numerical implementation, the initial conditions of the 'truth', as well as the priors assumed for the estimation, are given. The implementation of the EnKF used here is detailed in appendix D on page 159. The EnSRF implementation is according to [124]. The MCMC approach is the usual *Metropolis-Hastings* [261, 262, 158] using a 20% burn-in period and a thin chain[5].

4.5.1. Linear Bayesian Updating and Full Bayesian Updating

In this first numerical example the different PC updating approaches are compared to their sampling counterparts (EnKF and EnSRF) as well as to a full Bayesian update obtained by an MCMC approach on a complex scalar update.

The chosen prior is strongly skewed with PCE coefficients $x^\alpha = [1.4, 0.7, 0.2]$. The observation to be assimilated is 1.0. The associated

[5]Every 3rd sample of the posterior is used to reduce the correlation between samples.

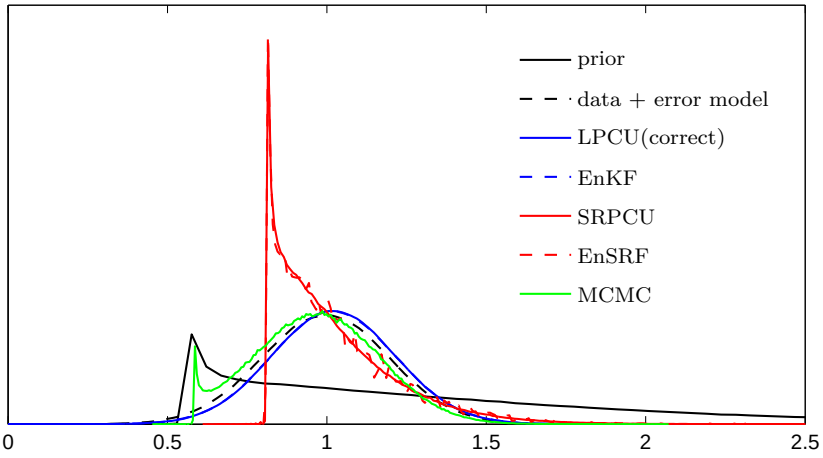

Figure 4.2.: Comparison of LPCU and SRPCU to EnKF, EnSRF, and MCMC on a scalar update. Shown are estimates obtained by a kernel density technique from either samples of the respective PCE or — in case of EnKF, EnSRF and MCMC — directly from the posterior. The sample size is in any case 10^6, except for EnSRF. There it is 10^4, due to computational limitations: a square matrix of that dimension has to be formed.

error model is chosen as $\mathcal{N}(0.0, 0.2)$, therefore $z(\omega) \sim \mathcal{N}(1.0, 0.2)$. The measurement operator is a direct, linear measurement with $y = x$ to avoid any influence from a complex model. This simplicity enables a clear demonstration of methodological differences and similarities. The posterior densities are estimated using a kernel density approach [72].

In Fig. 4.1, the three LPCU variants ('colinear', 'independent', and 'correct') are compared. As expected, the 'colinear' variant significantly overestimates the posterior variance, whereas the 'independent' variant underestimates it. The 'correct' variant produces a reasonable variance estimate — although this is barely visible due to the large errors of the other variants. See also table 4.1 for a numerical comparison for the first three moments.

Fig. 4.2 compares the 'correct' LPCU and SRPCU to their sampling-counterparts (EnKF for LPCU; EnSRF for SRPCU), as well as to the full Bayes update obtained by MCMC. The plot shows that the 'correct' LPCU result and the EnKF posterior are practically identical, while the SRPCU result is very similar to the EnSRF posterior; differences can be safely attributed to sampling noise. These similarities are to be expected, since LPCU/EnKF and SRPCU/EnSRF essentially implement the same two estimators — with the only difference being the different representations of the underlying RVs. Both square root posteriors are repositioned, rescaled versions of the prior. This is precisely what the Kalman filter — limited to Gaussian distributions — does. However, as discussed before, the exact form of the observation error model does *not* enter the square root updates of SRPCU and EnSRF; there, only the mean is corrected and the variance is reduced. On the other hand, the LPCU and EnKF posteriors are quite close to the observation error model, and a minor skewness is kept from the prior (see table 4.1). In these two methods, the exact form of the observation error model enters the method. However, note that all linear Bayesian methods (except the wrong implementations of LPCU) give the same mean and variance estimates. They only differ in the higher moments.

From Fig. 4.2, differences between the linear Bayesian methods and the fully non-linear Bayesian update are also evident: the MCMC posterior is a re-normalised multiplication of the prior and the likelihood, as stated by Bayes' theorem. This result cannot be expected from a linear update — except in the Gauss-linear case, where the Kalman filter does exactly this. However, the orders of magnitude faster computation of the linear

Functional	Mean	Variance	Skewness
Observation + Error Model	1.0001	0.0400	-0.0012
Prior	1.3969	0.5666	1.515 7
LPCU(colinear)	1.0279	0.9584	0.0789
LPCU(independent)	1.0263	0.0024	1.5070
LPCU(correct)	1.0266	0.0373	0.0267
EnKF	1.0259	0.0375	0.0257
SRPCU	1.0262	0.0374	1.5084
EnSRF	1.0266	0.0373	1.5049
MCMC	0.9739	0.0368	0.1721

Table 4.1.: Mean, variance, and skewness for SRPCU, the three LPCU variants, EnKF, EnSRF, and MCMC posteriors on a scalar update.

Bayesian posteriors compared to MCMC sampling has to be kept in mind. In table 4.1, one may see that all linear Bayesian approaches have the same mean and variance, and that these results are comparable to the MCMC result. Major differences may lie in higher moments. This is also where the square-root (SRPCU, EnSRF) and non-square-root (LPCU, EnKF) approaches differ. This makes it evident that they essentially are implementations of two different estimators which coincide for the Gaussian case. Among other results, this will be demonstrated in the next numerical example.

4.5.2. Application to a Gauss-Linear Model

The *Gaussian random walk* is a linear, scalar, discrete time model to demonstrate the SRPCU behaviour on a Gauss-linear system. It is compared to the two problematic LPCU variants, as well as EnKF and EnSRF. The 'correct' LPCU is left out, since in this sequential updating setting the growing of the PCE basis already poses a computational challenge.

The Gaussian random walker starts from some initial condition x_0 and at each time increment the random walker changes its position according to

$$x_{t+1} \quad = \quad x_t + a_t(\omega). \tag{4.20}$$

The random increment $a_t(\omega)$ follows a Gaussian distribution $\mathcal{N}(0, \sigma_a)$. Practically this would mean that the model is random. However, for each run the same random number generator with the same seed is used for advancing the model, so that it is no longer random and appears to follow some deterministic (but unspecified) evolution equation. Therefore the parameter σ_a is assumed to be known exactly in all experiments and the Gaussian random walk turns into a simple Gauss-linear model for which the original Kalman filter behaves optimally and can be used as reference.

4.5.2.1. Implementation

For the 'truth' run, at each t a sample of $a_t(\omega)$ is obtained from the model random generator and added to the model state.

The a priori state estimate $x_0(\omega)$ for this model is necessarily also Gaussian, with $x_0(\omega) \sim \mathcal{N}(0, \sigma_i)$. Its representation by an ensemble is quite obvious: a sample from the initial condition is obtained and at each t a sample of $a_t(\omega) \sim \mathcal{N}(0, \sigma_a)$ (generated independently from x_t) is added according to Eq. (4.20). Also the PCE representation is straightforward: for this model a PCE of order one is exact, as it can exactly represent a Gaussian distribution. The mean of the estimator does not change by the evolution (*cf.* Eq. (4.20)). The variance of the independent Gaussian RVs $x_t(\omega)$ and $a_t(\omega)$ has to add, and therefore the standard deviations add by the Pythagorean theorem

$$\mathrm{std}(x_{t+1})^2 = \mathrm{std}(x_t)^2 + \mathrm{std}(a_t)^2. \tag{4.21}$$

Since the first order terms of the PCE represent the standard deviations for a PCE of order one, this can be readily implemented.

4.5.2.2. Initial Conditions and Numerical Methods

In the following numerical examples, the 'truth' random walker is arbitrarily positioned at $x_0 = 123.76721401$ and the model standard deviation is $\sigma_a = 0.15$. The prior is $x_0(\omega) \sim \mathcal{N}(123, \sigma_i = 1.0)$. Noisy measurements are taken every five time steps, with a measurement error of

$\varepsilon(\omega) \sim \mathcal{N}(0, \sigma_n)$. Three experiments are performed using different measurement noise levels of $\sigma_n \in \{0.01, 0.1, 1.0\}$.

4.5.2.3. Results

Figs. 4.3–4.5 compare five different methods: the LPCU implementation suggested by Saad [307] (the 'colinear' observations), the implementation suggested by Zeng and Zhang [369] and Blanchard et al. [66] (the 'independent' observations), the SRPCU implementation (algorithm 2), a 'perturbed observations' implementation of the EnKF with a large ensemble of $N = 1000$ members, and EnSRF with $N = 1000$ ensemble members. The LPCU/SRPCU implementations use a PCE of order $P = 1$, since this is known to be sufficient for this task. All results are normalised with respect to the Kalman filter result, since it is known to be exact for Gauss-linear problems. The task is to estimate the scalar state from measurements with three different noise levels. Due to the extreme simplicity of the model, numerical and sampling effects can be compared efficiently.

In Fig. 4.3 (a) the relative error for the mean estimate $\mathbb{E}(\hat{x}) = \hat{\hat{x}}$ is shown (see appendix C.2). It is not surprising to see that differences between the methods are negligible. In Fig. 4.3 (b) variance estimates for the same methods are shown. There one would see some noise for the EnKF and EnSRF which is to be expected due to sampling errors. The overestimation of variance for the 'colinear' LPCU implementation is invisible, but the error has a similar magnitude as the sampling error of the two EnKF implementations. The complete removal of the additional variance in the 'independent' LPCU approach results in a strong underestimation of the variance. This completely hides the minor errors of the other methods. The *root mean square error* (RMSE, see appendix C.1), shown in Fig. 4.3 (c), is also practically identical for all methods except the 'independent' LPCU. Due to the slight but consistent overestimation of variance for the 'colinear' LPCU, also its RMSE is slightly larger (not visible). The SRPCU approach exactly reproduces the Kalman filter result in all cases, as expected (unfortunately invisible in the plots). Comparing Fig. 4.3 to Fig. 4.4, the picture starts to turn. Due to the larger measurement error, the variance estimates (b) of the 'colinear' LPCU approach become worse when compared to the EnKF implementations. Also the mean estimates (a) are no longer exactly reproducing exactly the KF results, resulting

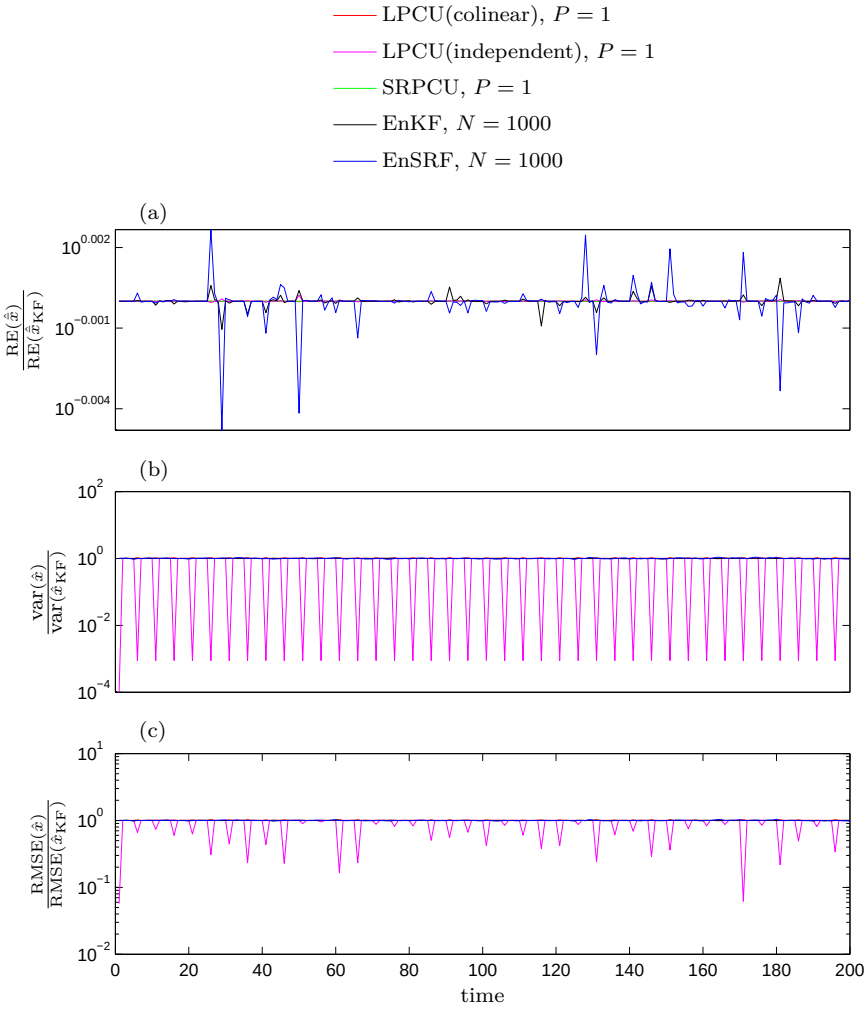

Figure 4.3.: Comparison of different methods when applied to the simple random walk model with a small measurement noise of $\sigma = 0.01$. Plot (a) compares the relative error of the mean, plot (b) compares variance estimates, and plot (c) compares the root mean square error for each of the estimators. All quantities are plotted relative to the respective Kalman filter solution.

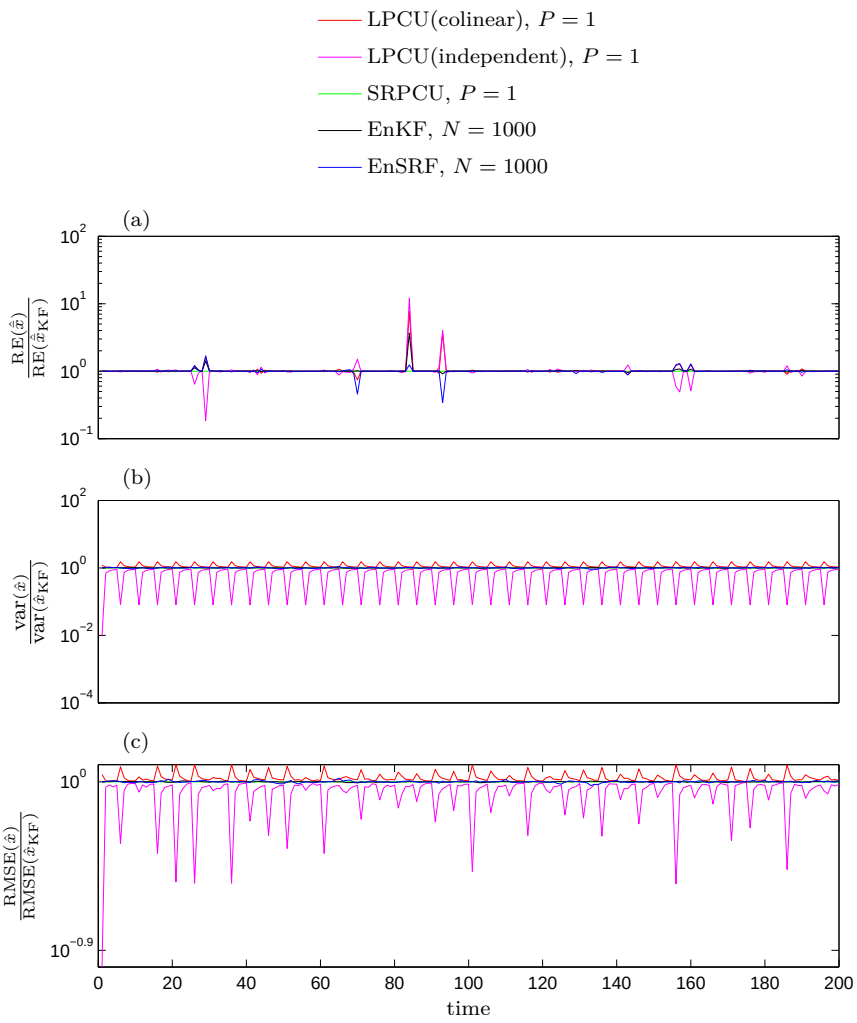

Figure 4.4.: Comparison of different methods when applied to the simple random walk model with a medium measurement noise of $\sigma = 0.1$.

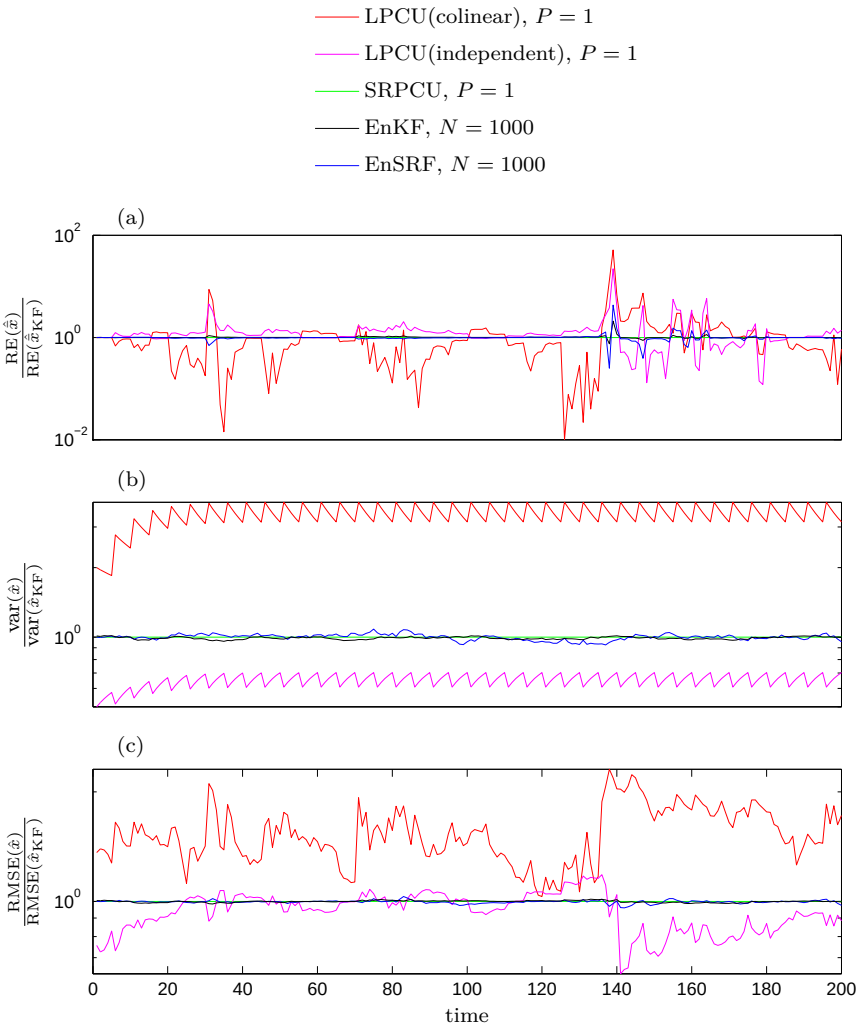

Figure 4.5.: Comparison of different methods when applied to the simple random walk model with a large measurement noise of $\sigma = 1$.

in an overall larger RMSE (c). The 'independent' LPCU approach improves over the previous example, but is still severely underestimating the variance. The SRPCU remains exact. In Fig. 4.5 the results of the two LPCU implementations are quite problematic. The measurement error now has a magnitude similar to the forecast uncertainties. For the 'colinear' LPCU approach, this leads to an almost exact cancellation of the variance when computing the innovations $z(\omega) - y(\omega)$. The 'independent' LPCU approach improves again, but still underestimates the variance. The problems of the two LPCU implementations appear because the variances of the two RVs $z(\omega)$ and $y(\omega)$ should *add*, since they are by definition uncorrelated — and it does not for both of them. This leads to the severe errors observed in Fig. 4.5. However, the SRPCU, as well as the ensemble Kalman variants (up to sampling errors), reproduce the Kalman filter results.

Therefore we conclude that the 'colinear' LPCU implementation may be acceptable when the observation noise level is small compared to the forecast. The 'independent' LPCU approach is problematic for all three examples. SRPCU is exactly reproducing the Kalman filter results. This is something one would expect — given that the original Kalman filter is the low-order part of LPCU [8] and SRPCU.

4.5.3. Application to Lorenz-63

In this series of experiments SRPCU is compared to EnKF and EnSRF on a much more complex, non-linear model: the *Lorenz-63* equations. This has been a popular numerical example for filter performance comparison for quite some time; therefore we choose to use it here, too.

The state evolution of the Lorenz-63 model, $\dot{\boldsymbol{u}} = \frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t} = f(\boldsymbol{u})$; $\boldsymbol{u}(0) = \boldsymbol{u}_0$, is described by the following set of ordinary differential equations (ODEs):

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= s(y - x) \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= rx - y - xz \\ \frac{\mathrm{d}z}{\mathrm{d}t} &= xy - bz, \end{aligned} \tag{4.22}$$

with three parameters s, r and b. The Lorenz-63 model shows chaotic

behaviour and is very sensitive to the initial conditions. For this reason we model them as independent Gaussian RVs:

$$\begin{aligned} x_0(\omega) &\sim \mathcal{N}(x_0, \sigma_1) \\ y_0(\omega) &\sim \mathcal{N}(y_0, \sigma_2) \\ z_0(\omega) &\sim \mathcal{N}(z_0, \sigma_3). \end{aligned} \tag{4.23}$$

In a second experiment, also the parameters are considered as uncertain. There,

$$\begin{aligned} s(\omega) &\sim \mathcal{N}(s_0, \sigma_1) \\ r(\omega) &\sim \mathcal{N}(r_0, \sigma_2) \\ b(\omega) &\sim \mathcal{N}(b_0, \sigma_3) \end{aligned} \tag{4.24}$$

are used. Due to the appearance of these RVs, the deterministic model Eq. (4.22) turns into a system of stochastic differential equations (SDEs, *e.g.* [282]),

$$\begin{aligned} \frac{\mathrm{d}x(\omega)}{\mathrm{d}t} &= s(\omega)(y(\omega) - x(\omega)) \\ \frac{\mathrm{d}y(\omega)}{\mathrm{d}t} &= r(\omega)x(\omega) - y(\omega) - x(\omega)z(\omega) \\ \frac{\mathrm{d}z(\omega)}{\mathrm{d}t} &= x(\omega)y(\omega) - b(\omega)z(\omega), \end{aligned} \tag{4.25}$$

which need to be integrated in time to obtain the evolution of the stochastic state vector $\boldsymbol{u}(\omega) = (x(\omega), y(\omega), z(\omega))^T$.

4.5.3.1. Implementation

For sampling approaches — such as the EnSRF and EnKF — the initial conditions are sampled according to Eq. (4.23), and the parameters according to Eq. (4.24). Each sample can be integrated forward in time according to Eq. (4.22).

For the PCE-based approach discussed in this work it is suitable to directly use the truncated PCE Eq. (3.4) as representation for the involved RVs (an approach also followed in [321, 6] for the *Lorenz-84* model): the state RVs $x(\omega), y(\omega)$, and $z(\omega)$ are replaced by the Hermite transforms (see

appendix B.4) $\xi := \mathscr{H}(x)$, $\eta := \mathscr{H}(y)$, $\zeta := \mathscr{H}(z)$. The transform is truncated by projection on the finite subspace generated by $\{H_\alpha \,|\, \alpha \in \mathcal{J}_Z\}$, resulting in $\widehat{\xi}(\omega)$, $\widehat{\eta}(\omega)$, and $\widehat{\zeta}(\omega)$. Here, $\widehat{\ }$ denotes the projection operation. The parameter RVs $s(\omega)$, $r(\omega)$ and $b(\omega)$ are treated similarly, resulting in $\widehat{\sigma}$, $\widehat{\rho}$, and $\widehat{\beta}$. For simplicity the same subspace is used for all quantities.

However, due to the truncation Eq. (4.25) cannot be satisfied exactly anymore: for example the result of a product of two truncated PCEs does not necessarily lie in that subspace anymore. To solve this problem we perform a Galerkin projection back onto that subspace. The final result is then the stochastic evolution Eq. (4.25) projected onto the subspace:

$$\begin{aligned} \frac{\mathrm{d}\widehat{\xi}}{\mathrm{d}t} &= \widehat{\mathsf{Q}}_2(\widehat{\sigma}, \widehat{\eta} - \widehat{\xi}) \\ \frac{\mathrm{d}\widehat{\eta}}{\mathrm{d}t} &= \widehat{\mathsf{Q}}_2(\widehat{\rho}, \widehat{\xi}) - \widehat{\eta} - \widehat{\mathsf{Q}}_2(\widehat{\xi}, \widehat{\zeta}) \\ \frac{\mathrm{d}\widehat{\zeta}}{\mathrm{d}t} &= \widehat{\mathsf{Q}}_2(\widehat{\xi}, \widehat{\eta}) - \widehat{\mathsf{Q}}_2(\widehat{\beta}, \widehat{\zeta}). \end{aligned} \tag{4.26}$$

The terms $\widehat{\mathsf{Q}}_2(\cdot,\cdot)$ denote the truncated/projected Hermite transform of the product of two RVs, which may be computed analytically from the PCEs of the RVs (see appendix B.4 for details).

The time integration method used for both the ensemble-based approach as well as the PCE implementation is a variable order *Adams-Bashforth-Moulton PECE* solver [79]. The accuracy of the method is set to very high levels, such that modelling errors are negligible.

4.5.3.2. Initial Conditions and Numerical Methods

In a first task, the data is simulated by linear measurements of the state variables ($h(\cdot) = \boldsymbol{H} = \boldsymbol{I}_3$) and adding samples of zero-mean Gaussian noise with known covariance $\boldsymbol{C}_{\boldsymbol{\varepsilon}} = \sigma^2 \boldsymbol{I}_3$

$$\boldsymbol{z}(\omega) = \boldsymbol{I}_3 \breve{\boldsymbol{u}} + \mathcal{N}(0, \boldsymbol{C}_{\boldsymbol{\varepsilon}}). \tag{4.27}$$

The measurement standard deviation is chosen to be $\sigma = 3$ for the state-only estimation task. For the combined parameter-state estimation task,

$\sigma = 0.5$ and

$$\boldsymbol{H} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}, \tag{4.28}$$

meaning that only the state is observed. Therefore, the parameters have to be identified indirectly.

In a second task, a non-linear measurement operator is used:

$$\boldsymbol{z}(\omega) = \boldsymbol{u}(\omega)^T \boldsymbol{u}(\omega) + \mathcal{N}(0, \boldsymbol{C}_{\boldsymbol{\varepsilon}}). \tag{4.29}$$

In this case, the standard deviation of the additive measurement noise is chosen to be $\sigma = 2$. Note that the parameters are again unobserved and have to be identified indirectly.

In all cases, measurements are obtained once every time unit. As initial condition for the 'truth' state for all experiments we use [263, 123]

$$\breve{\boldsymbol{u}}_0 = (1.508870, -1.531271, 25.46091)^T. \tag{4.30}$$

The system is advanced in time using the classical choice of parameters [242]:

$$\breve{s} = 10, \quad \breve{r} = 28, \quad \breve{b} = 8/3. \tag{4.31}$$

For the state estimation tasks, the first guess for the state is — assuming lack of better knowledge — given by a normal distribution centred around a reasonable first guess for each component

$$\begin{aligned} x_0(\omega) &\sim \mathcal{N}(3, 1), \\ y_0(\omega) &\sim \mathcal{N}(-3, 1), \\ z_0(\omega) &\sim \mathcal{N}(20, 1). \end{aligned} \tag{4.32}$$

For the combined parameter-state estimation tasks, the first guess of parameters is chosen as

$$\begin{aligned} s(\omega) &\sim \mathcal{N}(11, 2), \\ r(\omega) &\sim \mathcal{N}(25, 4), \\ b(\omega) &\sim \mathcal{N}(9/3, 0.5), \end{aligned} \tag{4.33}$$

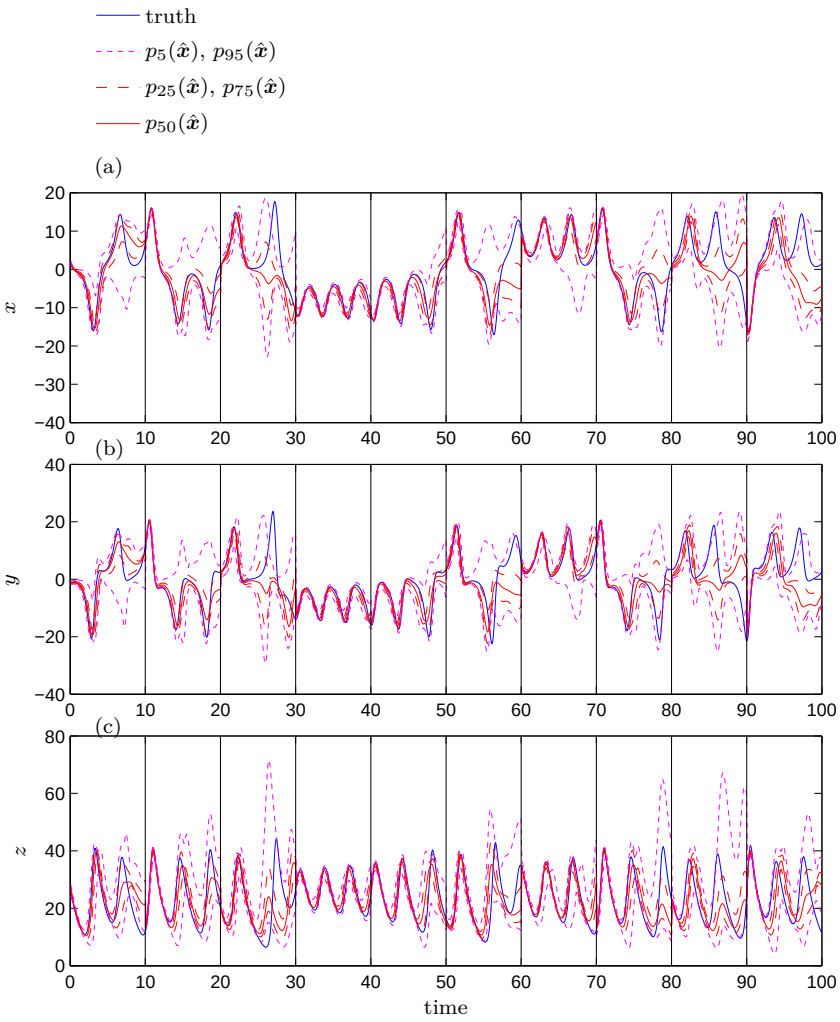

Figure 4.6.: Example of SRPCU with polynomial order $P = 3$ applied to Lorenz-63. Updates with measurements obtained according to Eq. (4.27) are performed once every 10 time units and are marked with a vertical bar. Shown are — for all dimensions x, y, z of the model — the 'truth', and some percentiles estimated from the updated PCE using sampling ($p_5, p_{25}, p_{50}, p_{75}$ and p_{95}).

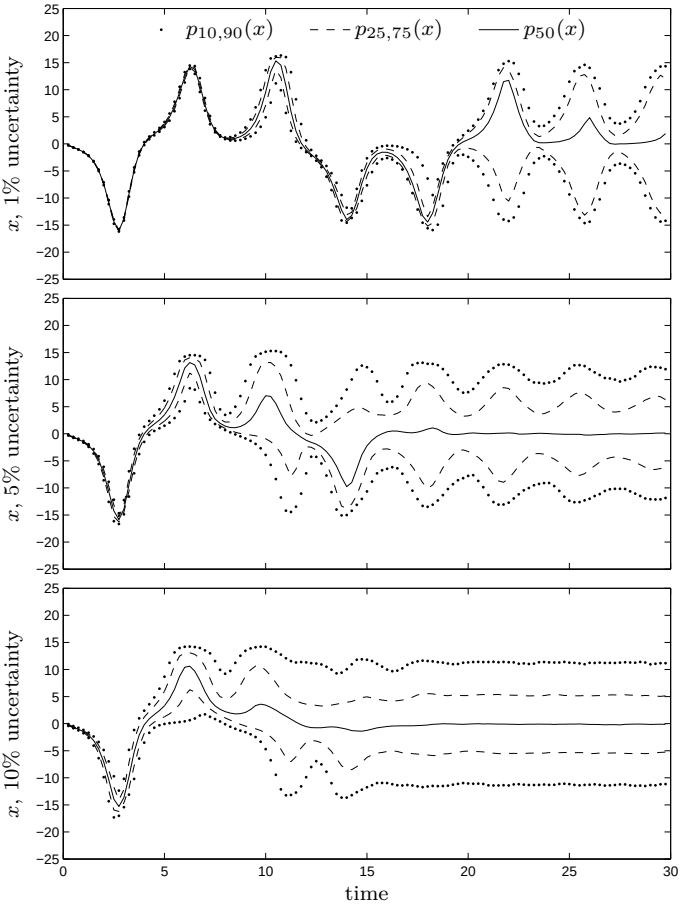

Figure 4.7.: Growth of different initial parametric uncertainties (1%, 5%, 10%) in $s(\omega)$, $r(\omega)$ and $b(\omega)$, starting from a practically negligible initial uncertainty in the state of $x_0(\omega), y_0(\omega), z_0(\omega) \sim \mathcal{N}(0, 10e-16)$. Some percentiles of the PDF of the x-component are plotted over time. The percentiles are estimated from 10000 random samples. The plot is similar for the y and z components.

while the first guess state remains as in Eq. (4.32). The mean parameter and state values clearly differ from the truth Eq. (4.31), but the first guess distribution 'covers' them sufficiently.

An example application of SRPCU for state estimation on Lorenz-63 is shown in Fig. 4.6. There, a long time interval of 10 is used in between updates to show the growing and updating of uncertainties (measurements are simulated according to Eq. (4.27)). Especially the time interval between 50 and 60 shows a strongly non-Gaussian distribution which is successfully updated. Due to the chaotic nature of the Lorenz-63 model, relatively small uncertainties in the parameters quickly cause major uncertainty on the state and a significant diversion from the 'truth' — this is demonstrated in Fig. 4.7.

4.5.3.3. Results: State Estimation and Reliability

In Figs. 4.8–4.20, the SRPCU approach is compared to EnSRF. Additionally, the same comparisons plots are shown for EnKF and EnSRF to demonstrate differences between the square root and non-square root estimator — remember that due to the growth of PCE basis it is impossible to implement the the non-square-root, 'correct' LPCU approach on such a sequential updating task in a straightforward way. A related publication is [8], where a state estimation and reliability analysis of the 'colinear' LPCU on the *Lorenz-84* model is performed.

The task is to estimate the state of the 'truth' run (*cf.* Fig. 4.8) from noisy data obtained by a linear state measurement. The following results are computed from 1000 repetitions of each experiment. For each repetition, the seeds of all random number generators are randomised: measurement noise, initial ensemble, and — in case of EnKF — observation ensemble. Linear, noisy data is assimilated once every timestep according to Eq. (4.27). Several different functionals $f(\cdot)$ of the estimates obtained from SRPCU, EnKF, and EnSRF are compared. Therefore, part (a) of each figure shows

$$\mathfrak{M}(f) = \frac{1}{N} \sum_{i=1}^{N} f(x_i), \tag{4.34}$$

the mean functional value over the 1000 repetitions. With the unbiased,

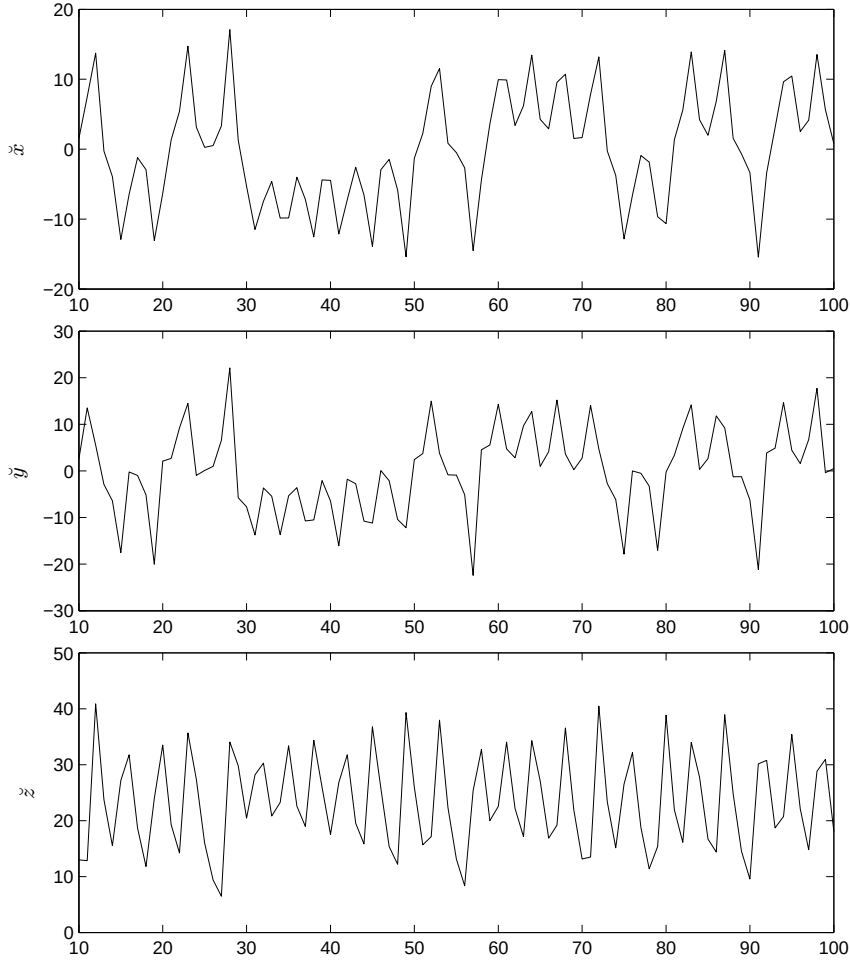

Figure 4.8.: Truth run of Lorenz-63 used for the experiment.

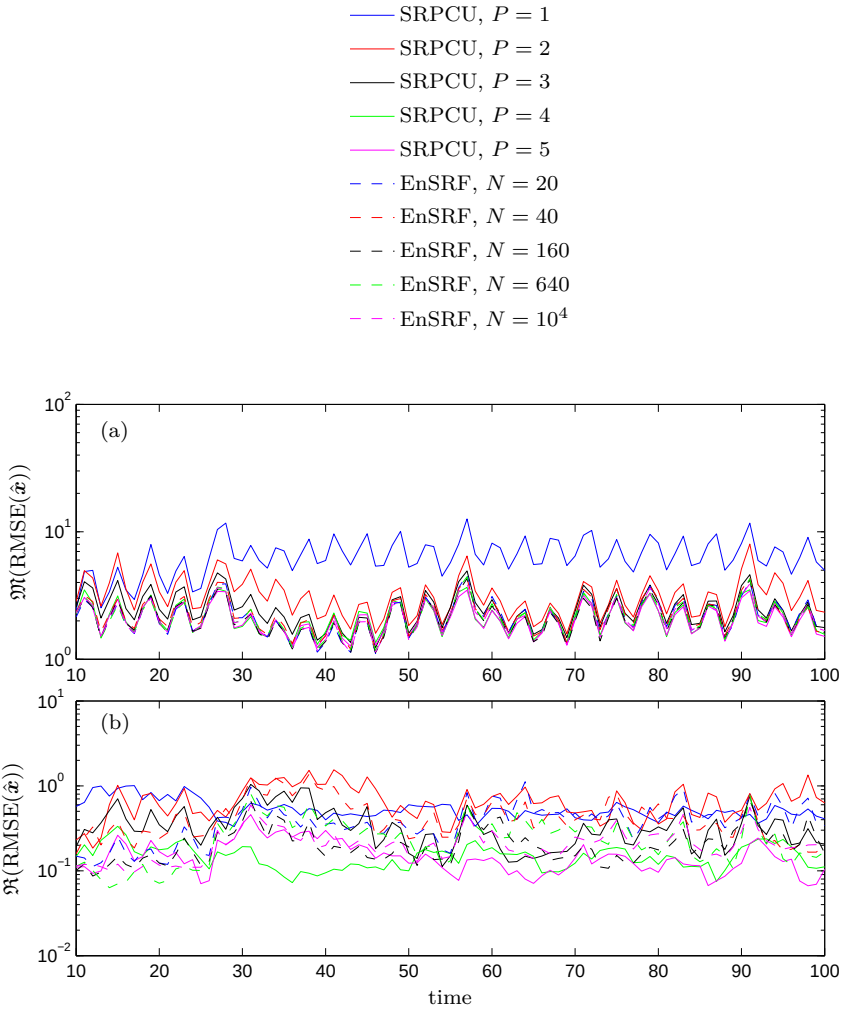

Figure 4.9.: RMSE and its reliability on Lorenz-63 (EnSRF, SRPCU).

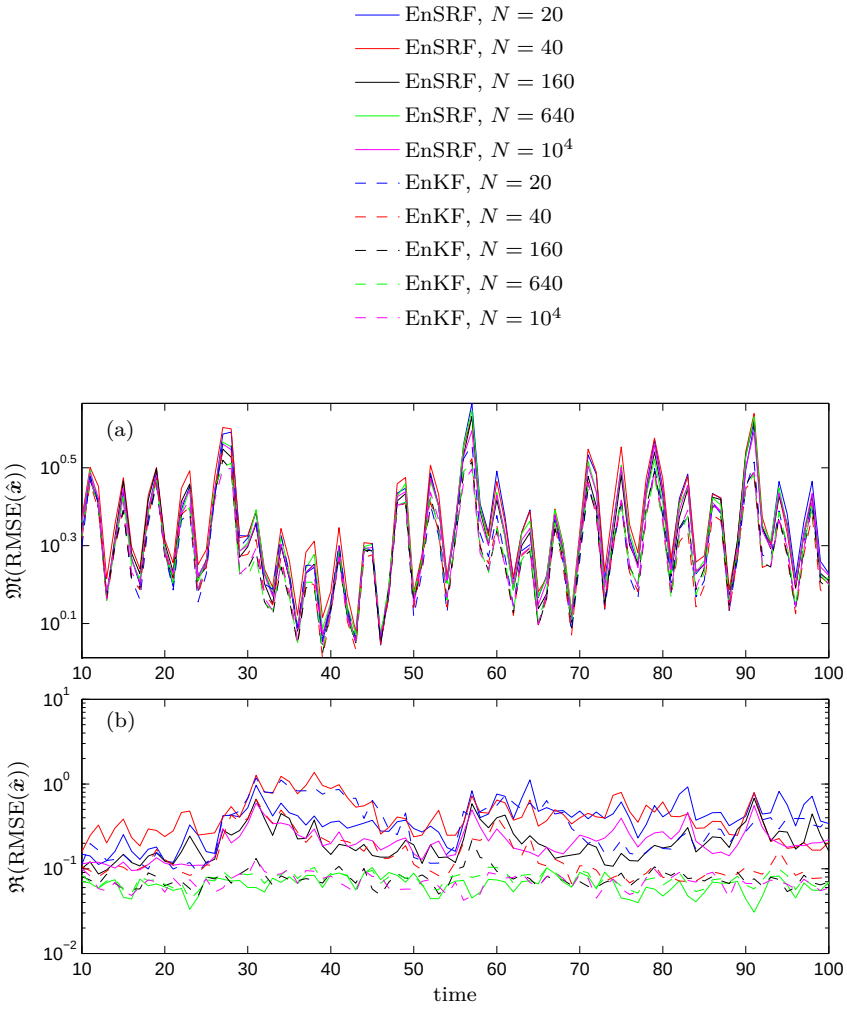

Figure 4.10.: RMSE and its reliability on Lorenz-63 (EnKF, EnSRF).

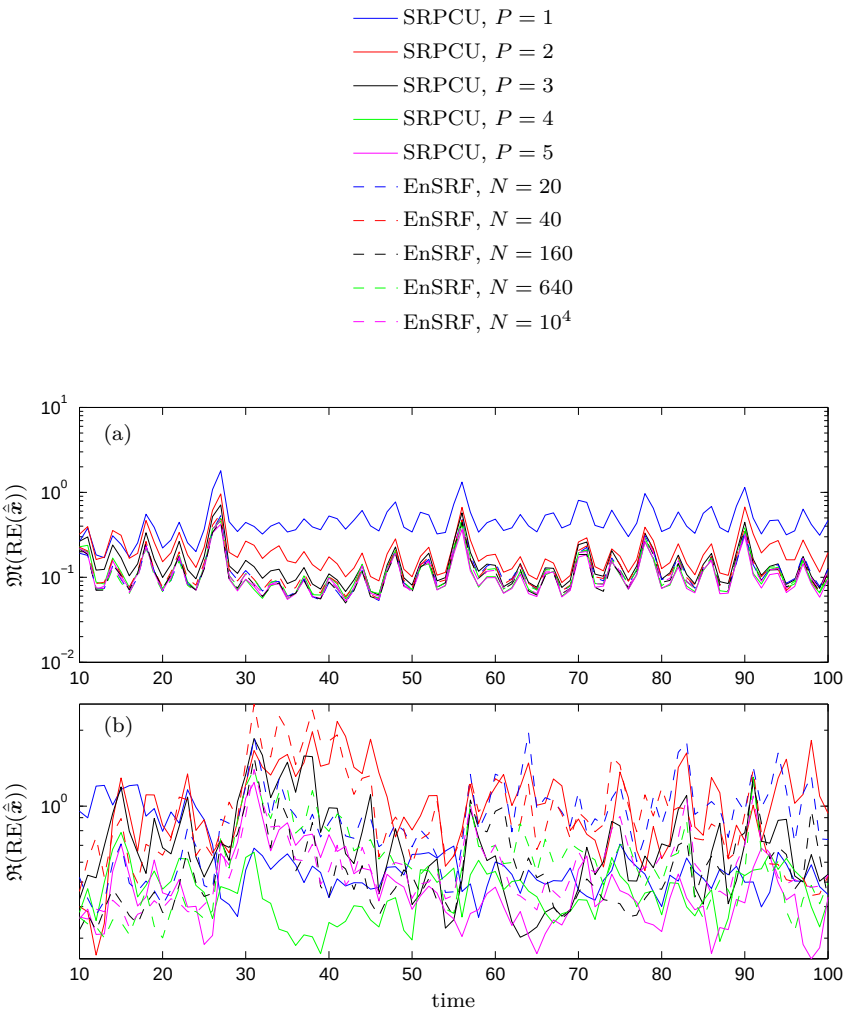

Figure 4.11.: Relative error of the mean and its reliability on Lorenz-63 (EnSRF, SRPCU).

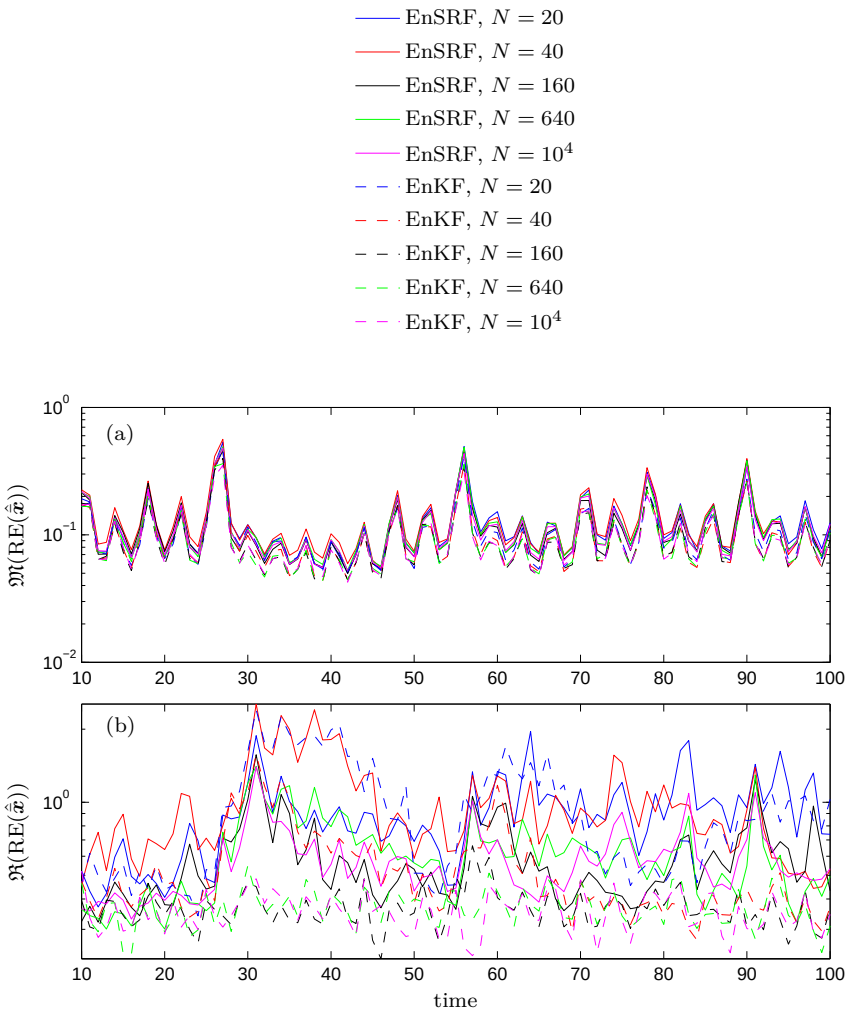

Figure 4.12.: Relative error of the mean and its reliability on Lorenz-63 (EnKF, EnSRF).

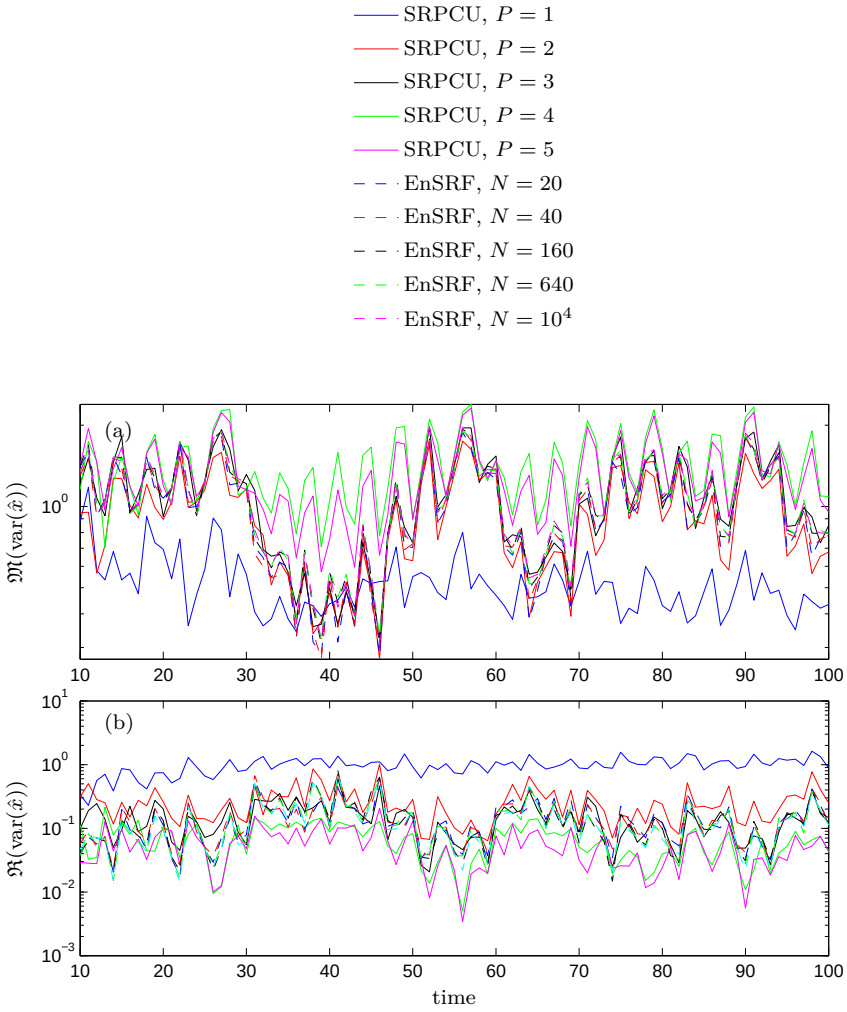

Figure 4.13.: Variance and its reliability on Lorenz-63 (EnSRF, SRPCU).

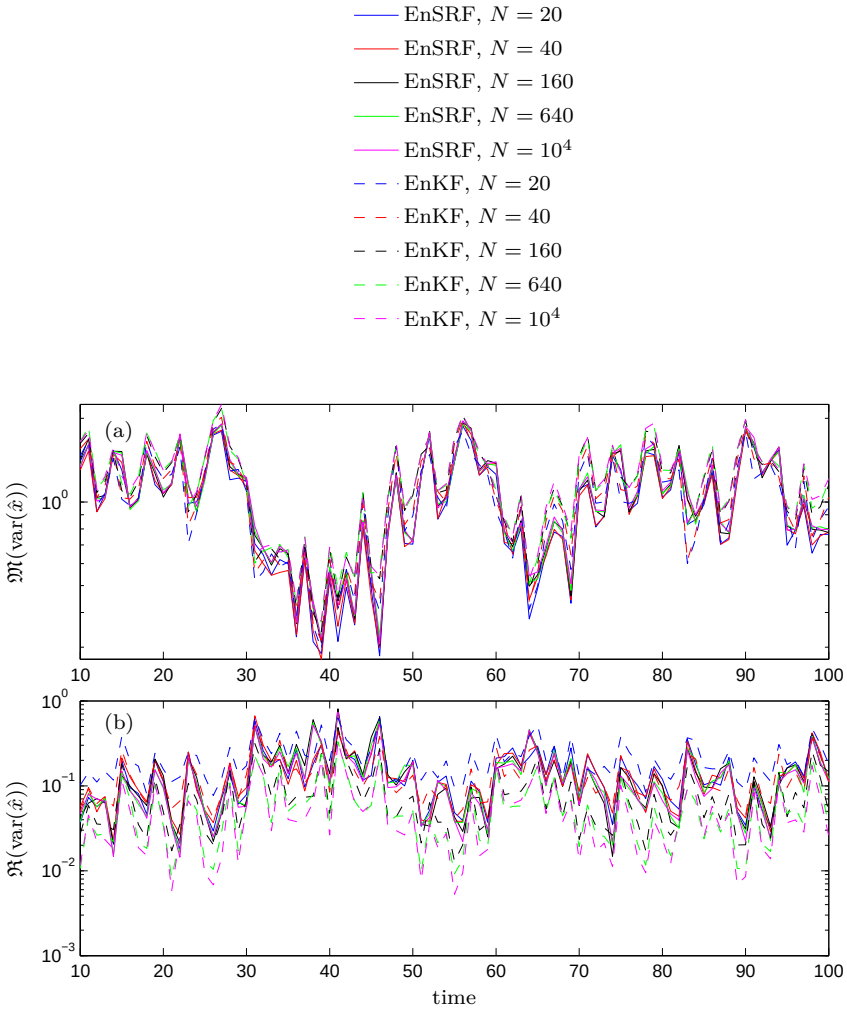

Figure 4.14.: Variance and its reliability on Lorenz-63 (EnKF, EnSRF).

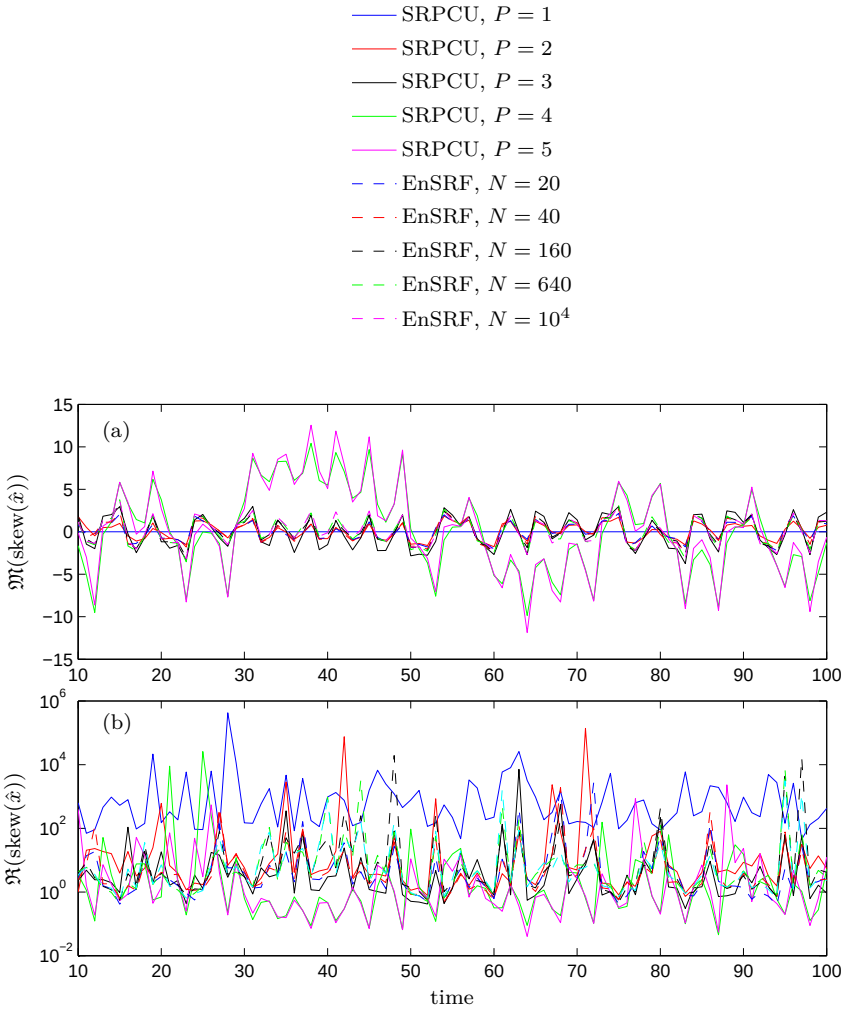

Figure 4.15.: Skewness and its reliability on Lorenz-63 (EnSRF, SRPCU).

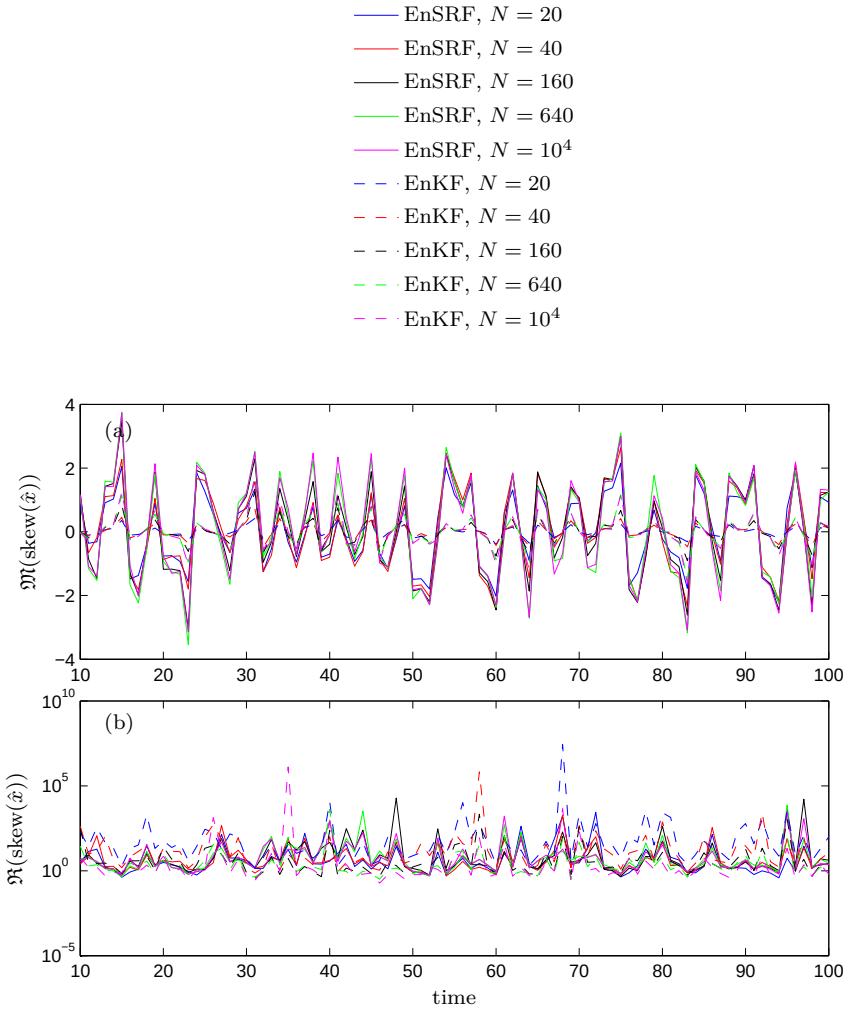

Figure 4.16.: Skewness and its reliability on Lorenz-63 (EnKF, EnSRF).

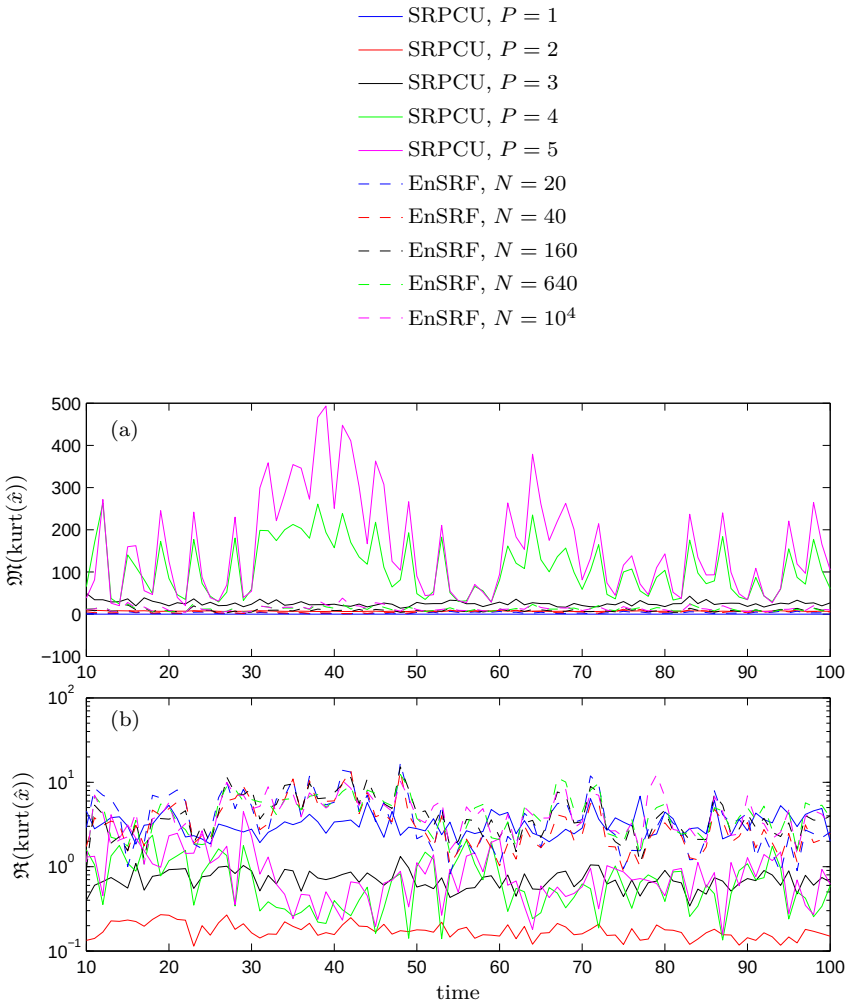

Figure 4.17.: Kurtosis and its reliability on Lorenz-63 (EnSRF, SRPCU).

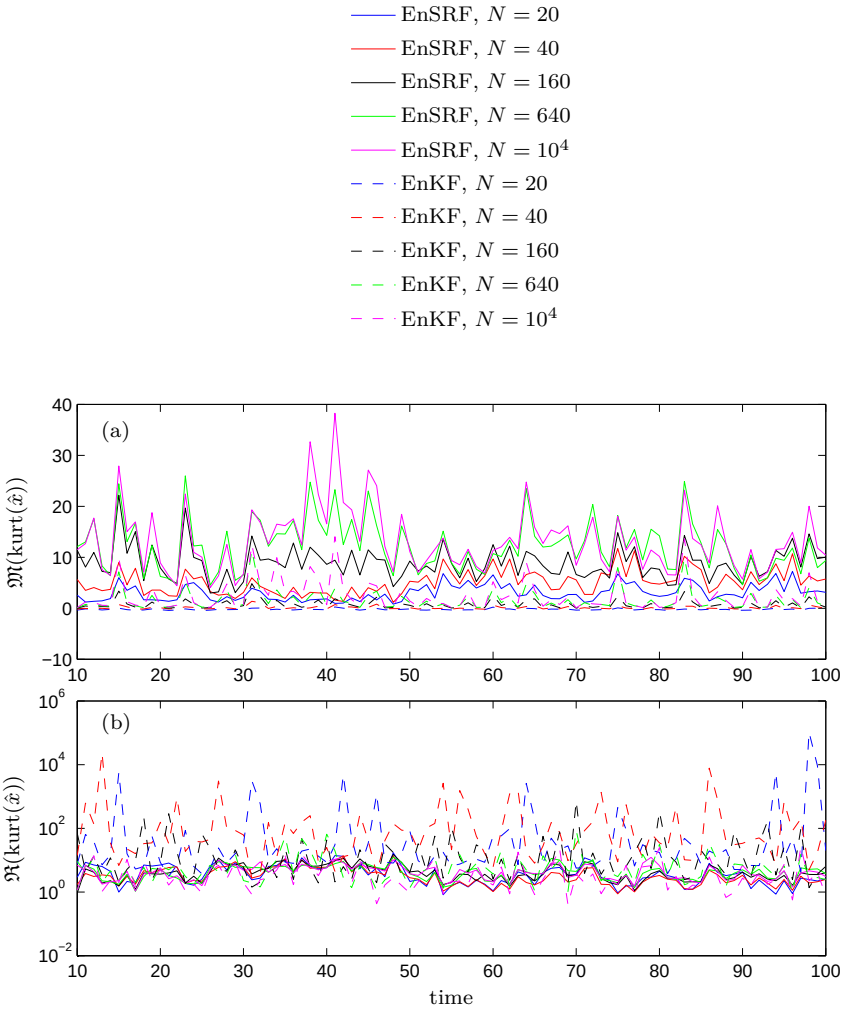

Figure 4.18.: Kurtosis and its reliability on Lorenz-63 (EnKF, EnSRF).

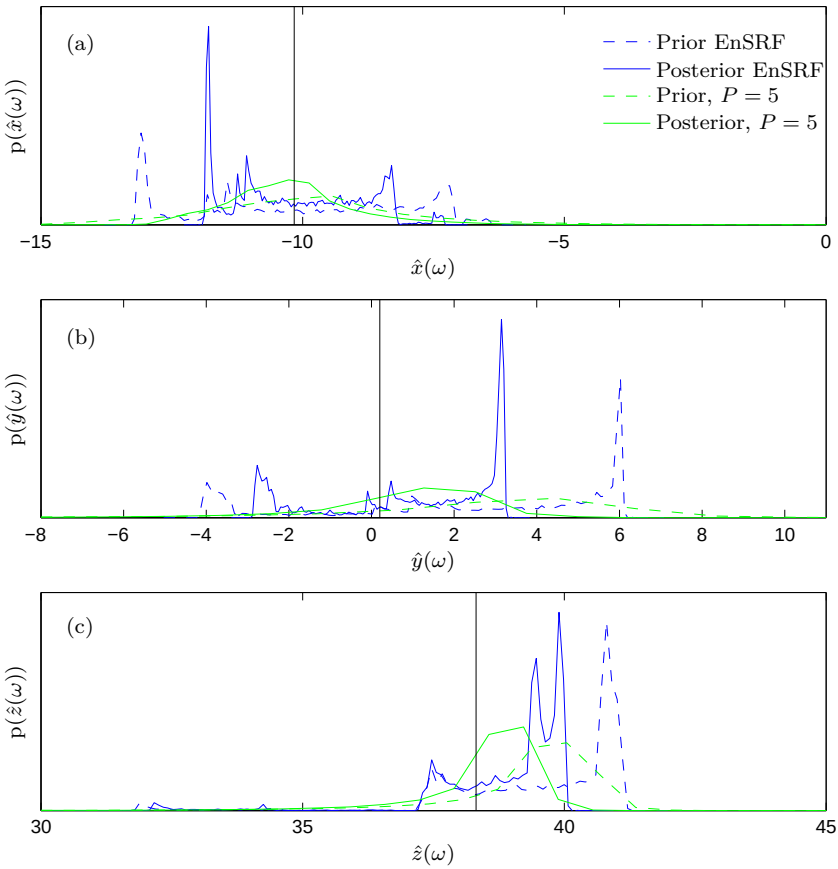

Figure 4.19.: Probability density estimates for SRPCU and EnSRF (with $N = 10^4$ members) on Lorenz-63 at $t = 80$ for all three components. The prior and posterior are shown for comparison. The vertical black bar marks the truth.

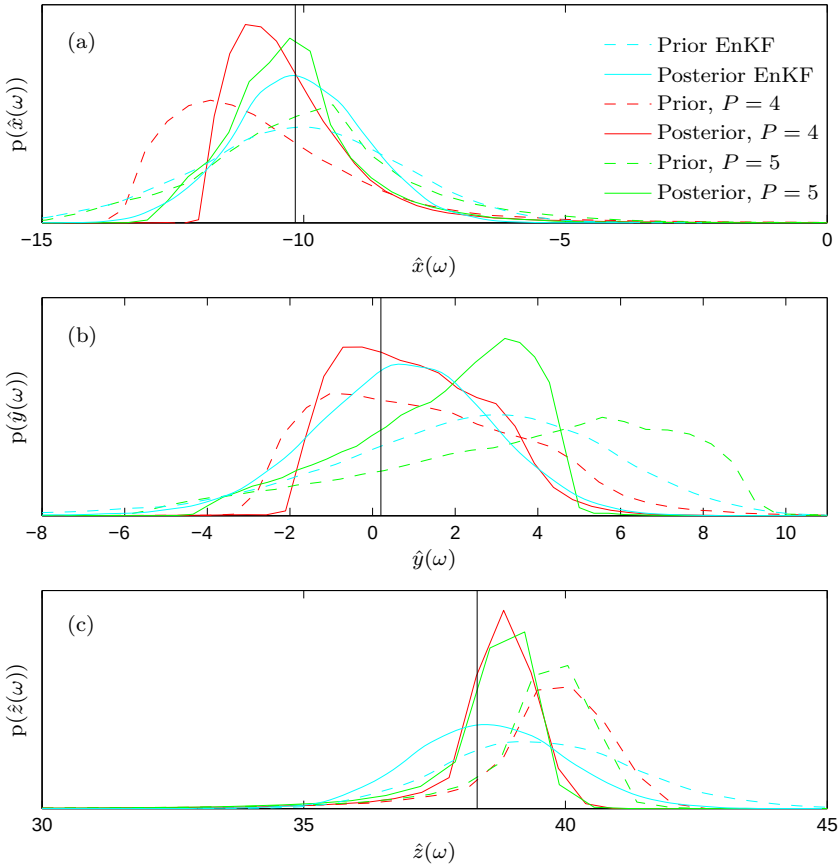

Figure 4.20.: Probability density estimates for SRPCU and EnKF (with $N = 10^4$ members) on Lorenz-63 at $t = 80$ for all three components. The prior and posterior are shown for comparison. The vertical black bar marks the truth.

sample estimated variance of the functional,

$$\mathfrak{V}(f) = \frac{1}{N-1}\sum_{i=1}^{N}(f(x_i) - \mathfrak{M}(f))^2, \tag{4.35}$$

one may obtain a measure of reliability of the respective functional $f(\cdot)$ via

$$\mathfrak{R}(f) = \mathfrak{V}(f)/\mathfrak{M}(f)^2, \tag{4.36}$$

the sample variance relative to the squared sample mean of the functional. Again, both are taken over the 1000 repetitions. $\mathfrak{R}(\cdot)$ is an indicator of the relative reliability of the functional value with respect to randomness in the estimation setups. A large value of the indicator means less reliability and vice versa. It is plotted in part (b) of each figure.

Fig. 4.9 shows the *root mean square error* (RMSE; see appendix C.1) results for SRPCU and EnSRF, and Fig. 4.10 shows EnKF and EnSRF. One immediately sees that for SRPCU the RMSE decreases with increased order, as one may expect. Good results start to be obtained with $P = 3$; a PCE with $P = 1$ is not enough. For the EnSRF, all presented ensemble sizes give similar RMSE results — but with different reliability. The EnSRF reliability is slightly worse than SRPCU, and SRPCU with $P = 4, 5$ clearly has the best, even when compared to very large ensemble sizes. Interestingly, the reliability of $P = 5$ is *worse* at certain times (*e.g.* $t \in [27, 50]$) than the one obtained with $P = 4$. This already hints at an especially 'difficult' regime of Lorenz-63. We will come back to this later. The RMSE for EnKF and EnSRF of all ensemble sizes is very similar. However, the reliability of the EnSRF estimates is lower at certain times. The reliability of the EnKF seems to have converged with an ensemble size somewhere around 160.

Fig. 4.11 shows relative error of the mean (see appendix C.2) for SRPCU and EnSRF, and Fig. 4.12 shows EnKF and EnSRF. The relative error clearly also improves with the order of SRPCU, and is already good for EnSRF with all ensemble sizes. Slight advantages for SRPCU over all EnSRF with $P = 4, 5$ may be observed, and in the phase of $t \in [27, 50]$, SRPCU with $P = 4, 5$ is clearly more reliable than EnSRF. Again, the reliability of the EnKF seems to have converged with an ensemble size somewhere around 160, and the behaviour of the reliability of EnKF and EnSRF is similar to the one observed for RMSE.

Fig. 4.13 shows variance estimates for SRPCU and EnSRF, and Fig. 4.14 compares EnKF to EnSRF. There, one may see that the variance estimate of SRPCU with $P = 1$ is unreliable. Its expectation is too low, and the reliability is low (indicated by high values in part (b) of the figures). SRPCU with $P = 2, 3$ behave similarly to EnSRF and EnKF of all orders. However, the variance estimate of SRPCU with $P = 4, 5$ is quite different — but with a higher reliability than all other methods and orders (except maybe for the highest orders of EnKF). This is again tied to the phase of $t \in [27, 50]$ (*cf.* Figs. 4.9 and 4.11). When looking at the truth run Fig. 4.8 on page 92, one may see that for $t \in [27, 50]$, the x and y components are both negative, indicating that the model is bound to one of the two attractors. However, the tendency to divert to the other attractor is apparently not negligible, and SRPCU is able to better represent the uncertainty in this phase by higher variance. For further investigation let us compare estimates of higher moments.

Fig. 4.15 shows skewness estimates for SRPCU and EnSRF, and Fig. 4.16 shows the same for EnKF and EnSRF. Note again that moments of the truncated PCE may be computed exactly (see appendix B.5). For the sampling-based approaches the usual sample skewness formula is used (see appendix C.3). In these figures it is immediately clear that the increased polynomial order directly translates to stronger skewness estimates. EnSRF tracks significantly less skewness than SRPCU; SRPCU already with $P = 2$ has similar skewness estimates than EnSRF even with large ensemble sizes. SRPCU with $P = 3$ outperforms all ensemble methods, and with $P = 4, 5$ has constantly different skewness estimates, also in the critical phase of $t \in [27, 50]$. The reliability of these estimates is also very high, as one may see from parts (b) of the two figures. EnKF practically tracks no skewness; only the two largest ensemble sizes slightly pick it up sometimes. This difficulty of the ensemble methods for higher moments becomes even more apparent when estimating kurtosis — as one may expect. This is demonstrated in Figs. 4.17 and 4.18. EnKF tracks almost no kurtosis, and the estimates are highly unreliable — except maybe for the largest ensemble. EnSRF is clearly better. On the other hand, SRPCU — especially with $P = 4, 5$ — behaves very differently.

This confirms the assumption about the significant tendency to switch attractors. The better representation of higher moments for SRPCU is due to the faster convergence of the PCE with respect to the L_2 norm. The significant result is that SRPCU is able to retain these properties for linear Bayesian updating in this example. However, it is strange to

see that EnSRF — while being a very similar square root method than SRPCU — also has problems with higher moments. To investigate this, Fig. 4.19 shows kernel density estimates at $t = 80$ for SRPCU and EnSRF (this time obtained from just a single run). The plot contains prior and posterior. The two cases are directly comparable, since the experimental setup is exactly the same. The estimate obtained for EnSRF is very noisy and seems to be multimodal. It also seems to contain outliers in the tails. While this may be partly due to the kernel estimation technique used here [72], the fact that an estimation grid of only 128 has been used counters this. Also the large ensemble size of $N = 10^4$ should allow for a smooth estimate. The estimates for SRPCU on the other hand are smooth and unimodal, as one would expect from this assimilation setup: the update frequency is high enough to stabilise the estimate on one mode of the complex state space of Lorenz-63. Possibly the observations of Wang et al. [354] and others [124], stating that ensemble square root schemes tend to produce collapsed ensembles with outliers representing the variance, is an explanation for this behaviour. Clearly this is not an issue for SRPCU, since the spectral decomposition is — and remains — orthogonal. Finally, Fig. 4.20 shows kernel density estimates at $t = 80$ for SRPCU and EnKF. Compared to SRPCU, the EnKF estimates are practically Gaussian, as expected from the previous results regarding higher moments. On the other hand, they are not collapsed and the mode of the posterior of EnKF is very well aligned with the truth. Note that a similar result may be achieved with EnSRF and *random rotations* [124] or, in fact, by SRPCU when leaving out the $\boldsymbol{V}^T$ in Eq. (4.15). Then, however, all information on higher moments is lost *by construction* and the posterior is multivariate Gaussian.

Though it remains unclear whether the EnSRF, EnKF, or the SRPCU results are better with respect to a fully non-linear Bayesian update, the square root results seem more conclusive with respect to the chaotic nature of Lorenz-63. Higher moments are much better represented by both EnSRF and SRPCU, while SRPCU by construction avoids the outlier problem and is able to represent the skewness and kurtosis better than EnSRF. This also translates to more realistic distribution estimates in phases where it is necessary, *i.e.* for $t \in [27, 50]$ in this example. Here, realistic is to be understood in the sense of 'being a distribution consistent with the underlying model'. The higher reliability of SRPCU compared to EnSRF is also to be expected: the EnSRF ensemble represents the complete state RV $\boldsymbol{u}(\omega)$ with all sample sizes — also the smallest ones. The PCE, on the other hand, needs a certain order of expansion to represent a large

enough part of the variance. However, the size of the EnSRF ensemble clearly influences the reliability, while for the SRPCU this influence does not exist. Therefore, reliability is generally better, given that the order of expansion is sufficiently high.

At this point the interested reader may ask why no comparison is made to a non-linear Bayesian update. Using MH sampling with a reasonable amount of samples, such an update over the full runtime of this small Lorenz-63 example is computationally already highly demanding. Additionally, the result would not allow us to compare the individual sequential updates. Usual particle filtering techniques using re-sampling are not applicable, since the Lorenz-63 model used here does not have a stochastic term, causing re-sampled particles to not depart from each other. An intermediate MCMC step for the re-sampling, and other enhancements and tweaks to tune the behaviour of a fully non-linear sequential Bayesian technique to Lorenz-63, are somewhat out of the scope of this work. Their results would also be questionable as strong reference cases due to the significant amount of careful tuning necessary to make these approaches work. A simple sequential combination of MH sampling and the PCE representation is not conclusive either: all approaches to obtain the posterior PCE necessarily contain an isoprobabilistic transform such as the *Nataf* or *Rosenblatt* transform (*e.g.* [225]), a multidimensional kernel density estimation technique (*e.g.* [316]) or other approximative methods and would not serve well as strong reference, too. Therefore, such a comparison would not add directly to the conclusions and is therefore omitted. Note that a comparison to full Bayes on a *single* update is performed section 4.5.1 on page 76.

4.5.3.4. Results: Combined State and Parameter Estimation with Linear Measurements

Figs. 4.21–4.24 compare the SRPCU approach to the EnSRF on an extended task: in addition to the initial conditions both methods shall estimate three uncertain parameters of a 'truth' run from linear, noisy measurements of the state. See also [9] for a related publication.

In the following, the focus is on parameter estimation quality. Only a single run is performed, but the observation noise random number generator has the same initial conditions for all runs. In Fig. 4.7 on page 90,

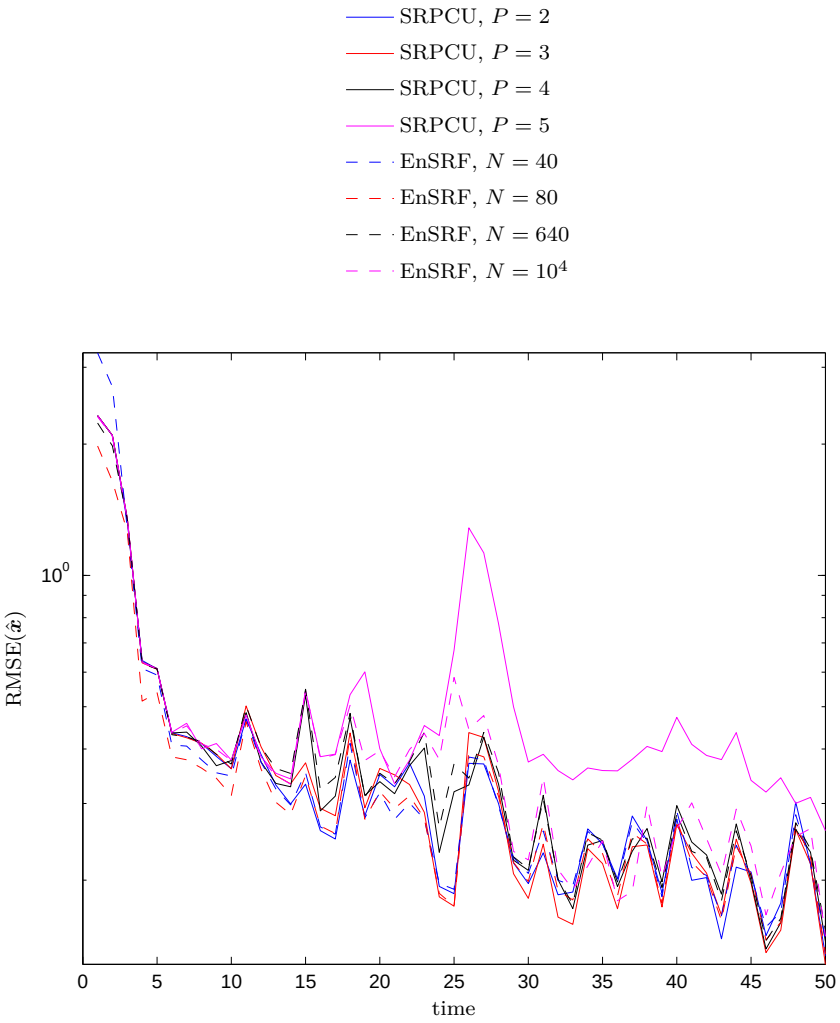

Figure 4.21.: Evolution of the overall RMSE for SRPCU and EnSRF on a combined parameter and state estimation problem.

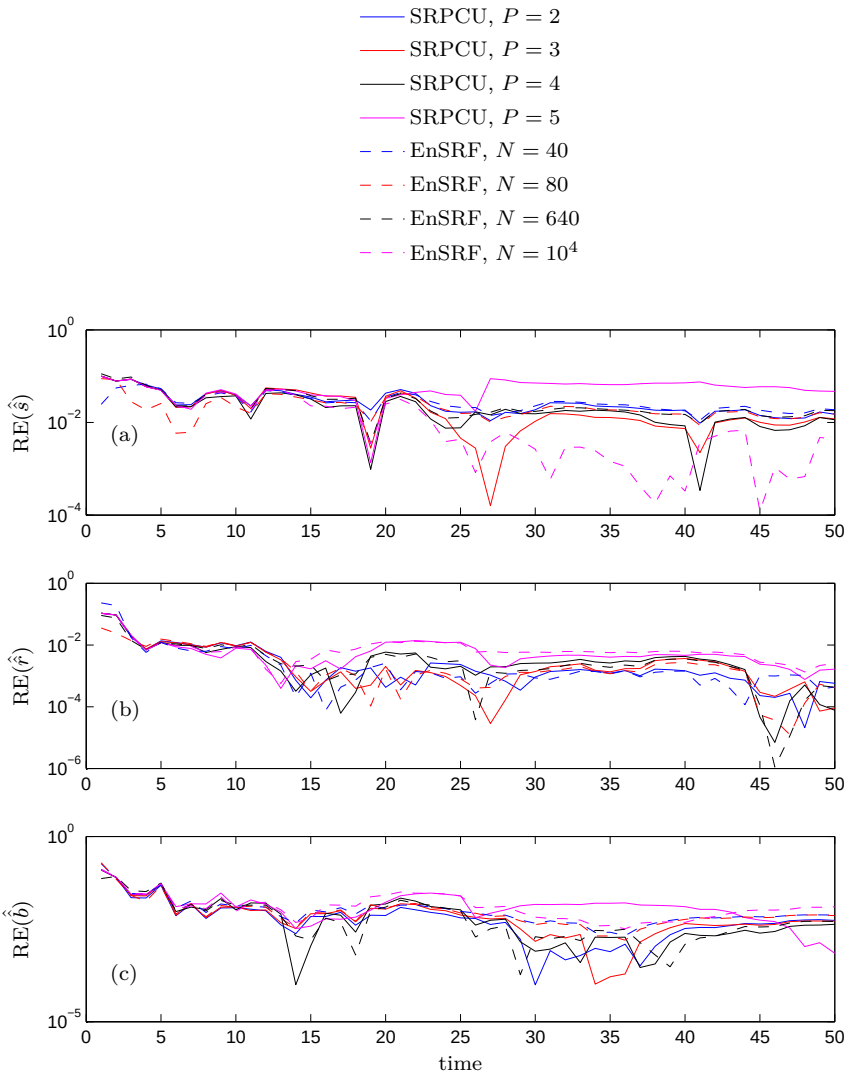

Figure 4.22.: Evolution of the relative errors of the mean for the parameter estimates.

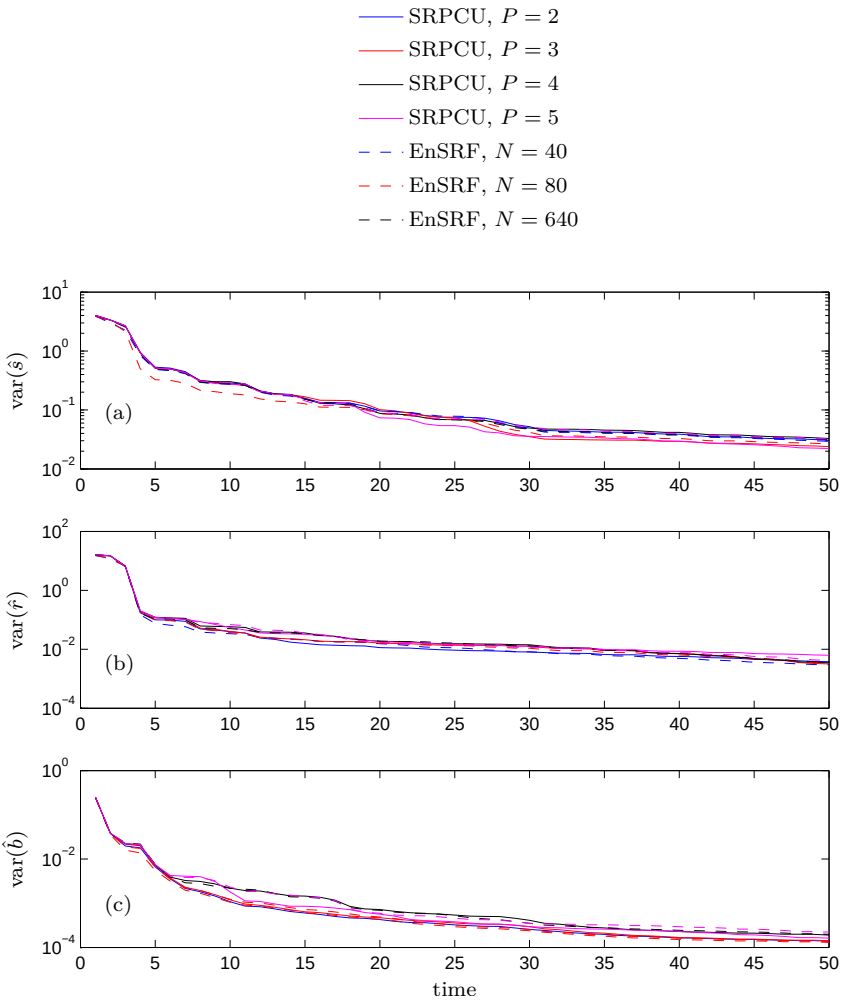

Figure 4.23.: Evolution of the estimated variance for the parameters, relative to the variance estimate obtained by EnSRF with $N = 10^4$ ensemble members.

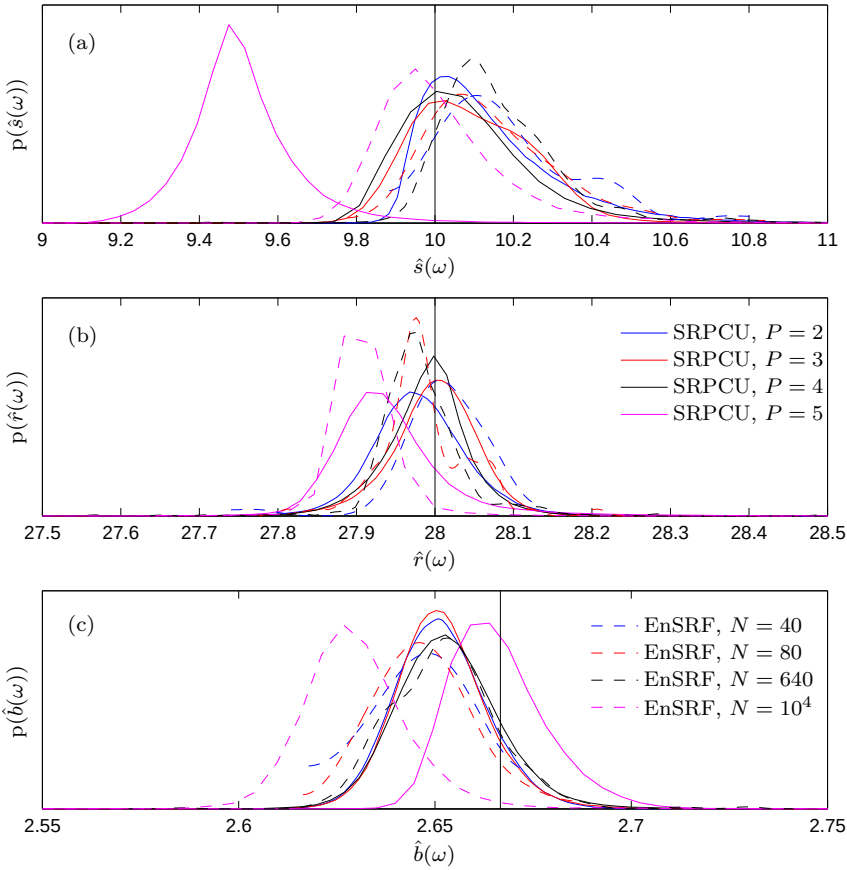

Figure 4.24.: Probability density estimates for $t = 50$ (after the last update) for SRPCU and EnSRF. The to-be-identified values $\breve{s}, \breve{r}$ and $\breve{b}$ are marked with a vertical bar.

the influence of different initial parametric uncertainties for Lorenz-63 is demonstrated. One can see that already a small amount of uncertainty in the parameters will quickly cause a significant uncertainty on the state, as expected from a chaotic system. Due to the quite non-linear relationship between state and parameters, and the still significant amount of observation noise, this task can be considered as difficult for any linear Bayesian scheme.

Fig. 4.21 shows that most methods produce a similar RMSE for this task. Interestingly, the higher order methods have a higher RMSE: EnSRF with $N = 10^4$ ensemble members and, obviously, SRPCU with $P = 5$. Fig. 4.22 shows relative errors in the mean for each method and for the three individual parameters. Interestingly, all methods behave very similar at times, whereas they differ quite strongly for other times. This may be due to the different regimes of the non-linear Lorenz-63 model, but also due to the different orders of the methods. For the first parameter, EnSRF with $N = 10^4$ seems to be the best method in this comparison, while for the other two it seems to be not so good. For the second, higher order SRPCU runs give the best results — but not the $P = 5$ one. On the other hand, for the last parameter SRPCU with $P = 5$ gives the best final result. Fig. 4.23 shows that the variance estimates for the three individual parameters are not very different for all methods. Finally, Fig. 4.24 shows probability density estimates for $t = 50$ (after the last update) for SRPCU and EnSRF and different orders. There one may see that the SRPCU ends up with quite close PDF estimates for the first two parameters, which also improve with improved order of the expansion — again with the exception of $P = 5$. Only the last parameter, b, seems to be difficult for all methods — again except SRPCU with $P = 5$, which obtains the best estimate. Strangely, the EnSRF results for $N = 10^4$ are clearly the worst. It can be speculated that the outlier problem (see previous experiment) causes problems for the parameter estimation as well.

From this task, one may conclude that the strongly non-linear connection between the parameters and the state variables poses difficulties for higher order methods. Since these methods are able to better represent the variance of the state variables (as demonstrated in the last experiment), the 'linearity conditions' given in Eq. (4.19) may be violated. This becomes especially apparent for SRPCU with $P = 5$, but also for EnSRF with $N = 10^4$. Therefore, this application is feasible but using very high order methods may not lead to the expected results.

It should be noted that we have also obtained results for a combined parameter-state estimation experiment with a reduced observation noise level of $\sigma = 0.1$. There, similar results are obtained as with the previous state-only estimation experiment and are therefore omitted. However, this confirms the suspicion that the amount of variance retained the estimation setup for above experiment is critical for linear Bayesian methods in general (see also the discussion at the end of section 4.4).

4.5.3.5. Results: Combined State and Parameter Estimation with Non-Linear Measurements

Figs. 4.25–4.29 compare the SRPCU approach to EnSRF on the same parameter estimation task, now even using a non-linear measurement operator. Already the RMSE comparison in Fig. 4.25 shows that the higher order methods perform more or less generally *worse*. This is true for both EnSRF and SRPCU. Therefore, a high-order EnKF run has been added in the RMSE comparison — but it shows similar behaviour. By looking at Figs. 4.26 and 4.27, one may see that the large RMSE is accompanied with larger errors in the mean as well as larger variance estimates. The variance estimates spread considerably more than with the linear state measurement in the previous experiment. Fig. 4.28 shows again probability density estimates for the posterior of the last time step. As expected from the RMSE and variance results, the higher order methods generally produce worse estimates for the parameters. The lower order methods, on the other hand, create quite good results, which is emphasised by Fig. 4.29, where the higher order methods have been left out. However, also there one may already see that SRPCU with $P = 4$ creates worse estimates than SRPCU with $P = 2, 3$.

From this experiment it becomes even more clear that a non-linear measurement operator must be treated with care — especially for higher order methods and in the presence of significant observation uncertainty. However, this has already been observed in the previous experiment for a linear — but indirect — measurement operator. In this case, a modified formulation of the update equation should be used that does not employ a linear measurement condition (*cf.* [359, p. 257f]). Alternatively, iterative improvements to the estimate could lead to better results (*cf.* [340, Eq. (3.51)]).

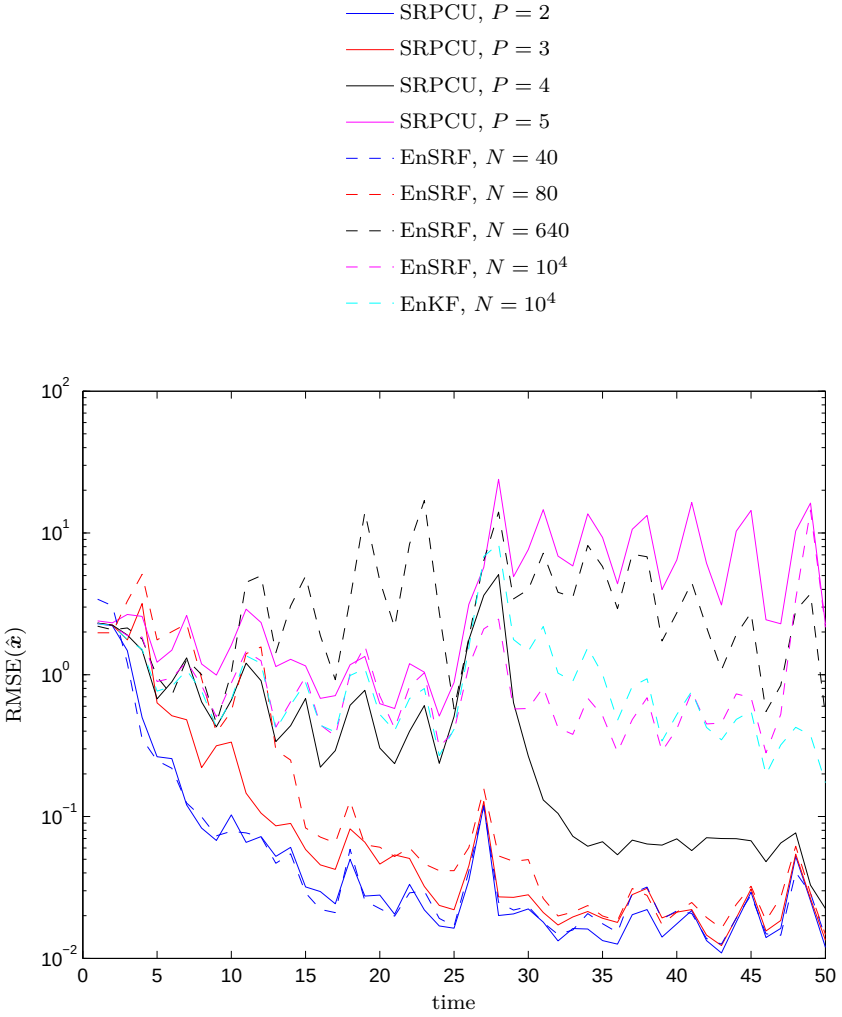

Figure 4.25.: Evolution of the overall RMSE for SRPCU and EnSRF on a combined parameter and state estimation problem with a non-linear measurement.

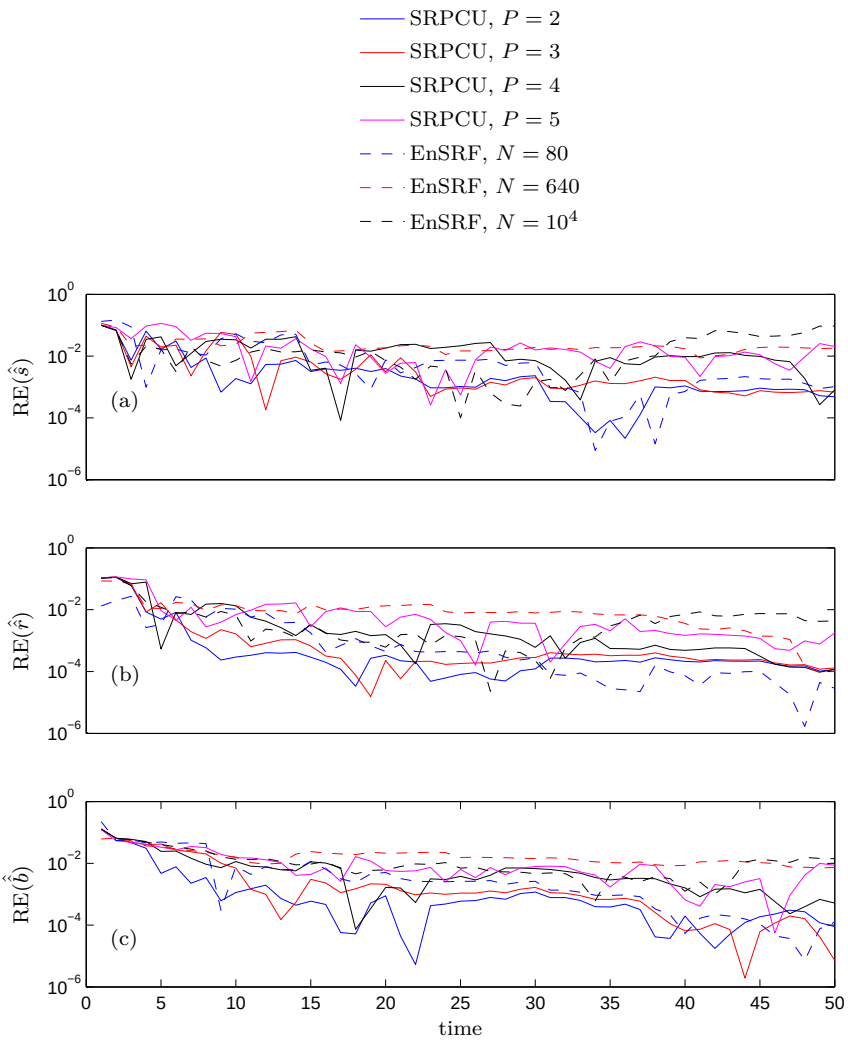

Figure 4.26.: Evolution of the relative errors of the mean for the parameter estimates with a non-linear measurement.

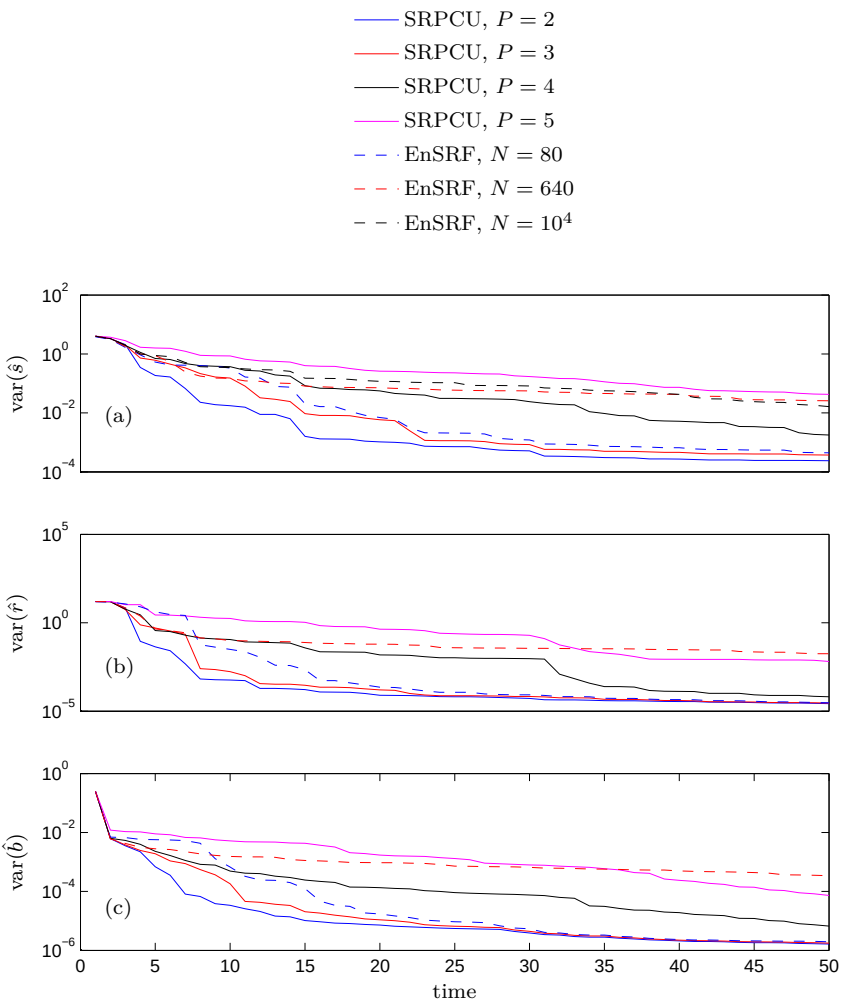

Figure 4.27.: Evolution of the estimated variance for the parameters with a non-linear measurement operator.

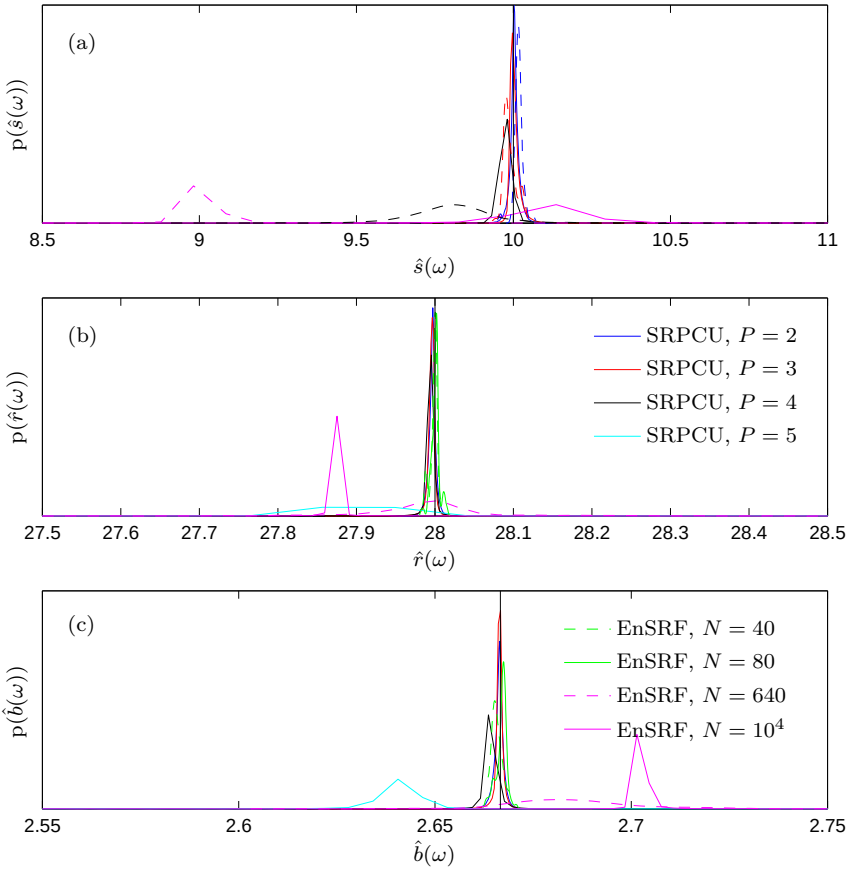

Figure 4.28.: Probability density estimates for $t = 50$ (after the last update) for SRPCU and EnSRF with a non-linear measurement operator. The to-be-identified values $\breve{s}, \breve{r}$ and $\breve{b}$ are marked with a vertical bar.

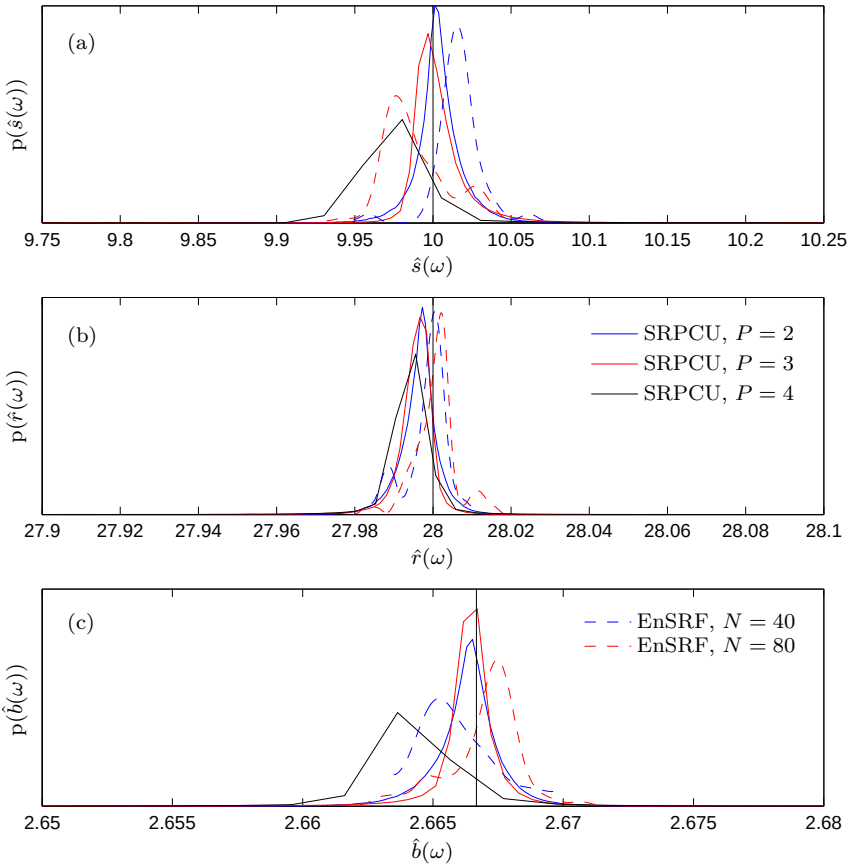

Figure 4.29.: Probability density estimates for $t = 50$ (after the last update) for low order SRPCU and EnSRF with a non-linear measurement operator. The to-be-identified values $\breve{s}, \breve{r}$ and $\breve{b}$ are marked with a vertical bar.

4.6. Summary of Numerical Examples

On a scalar example it has been demonstrated that all LCE methods create the same mean and variance results. Differences lie in higher moments, as demonstrated by the skewness: there, it is evident that square-root and non-square-root approaches are essentially two different estimators which coincide in the Gauss-linear case. This is also demonstrated by probability density estimates. While LPCU is the PCE-based equivalent to EnKF, SRPCU is similarly related to EnSRF. When compared to MCMC, all LCE methods result in different posterior distributions, as theory predicts: the subspace onto which the LCE projects is smaller compared to the full CE subspace.

SRPCU has been demonstrated to reproduce the Kalman filter result in case of Gauss-linear problems. For sequential estimation, the 'colinear' approximation to LPCU may be acceptable if the observation noise level is low compared to the forecast. Additionally, it is a conservative approximation in the sense that the posterior variance is higher than with the 'correct' implementation. The 'independent' approximation to the LPCU is more problematic, and it is an optimistic approximation (posterior variance is too low). SRPCU, on the other hand, does not involve any approximation (besides the linearity assumption inherent to all LCE approaches) and avoids the problem of a growing spectral basis in a consistent way.

On the more complex example of Lorenz-63, the difference between non-square-root and square-root approaches is also evident. For probability density estimates, for variance, and for higher moments in general, the SRPCU approach has been demonstrated to have advantages over EnSRF in several functionals — and their reliability with respect to the involved error terms and sources of noise. However, the precondition is that the order of PC expansion is sufficiently high — otherwise EnSRF may have advantages, albeit with low reliability if the ensemble size is low. Additionally, SRPCU does not have issues with maintaining the estimation subspace span, as does EnSRF: the probability density estimates do not start to be deteriorated by outliers after several updates. However, estimation tasks with non-linear relations and significant amounts of variance generally must be treated with care: it has been demonstrated that on these tasks, *all* LCE methods have difficulties — depending on the specific estimation setup, and especially on the involved observation

noise level. For such tasks, iterative schemes or other enhancements to the methods may be required.

The choice whether to use a non-square-root or square-root approach is certainly application dependent, but it is possible to give some indications. For the Lorenz-63 model it was found that the square-root approaches produce estimates which are more close to the prior — and thus more consistent with the actual model dynamics. The reason for this is that the distributional form of the measurement noise does not enter the update, only the first two moments. This, however, may not always be wanted — then one may have to use a non-square-root approach. However, there an approximative solution is necessary at the moment to avoid a growth of basis, and using a sampling representation may have advantages in that case. The choice between sampling and spectral represetation is otherwise indicated by what is available: if a spectral representation exists, and above mentioned properties of square-root approaches are suitable to the problem, SRPCU clearly should be used.

Chapter 5.

Multi-Scale Wavelet Analysis-based Localisation

The literature review in section 3.4.1.4 turns up a plethora of different regularisation approaches, especially for probabilistic LCE methods. It is clear that any application will require its own subset of these methods, together with possible modifications that further enhance performance. Over the course of this thesis, research has been conducted towards a method which aims at making use of evidence which cannot be precisely — but approximately — related to spatio-temporal quantities of the model.

The chapter is organised as follows: the presented approach is motivated by an example, and it is discussed why multiscale wavelet transform is thought to be an appropriate method. The ideas are connected to related work, and concluded with a short discussion.

5.1. Motivation

Common covariance localisation approaches for LCE methods consider the measurement — or 'source of evidence' — to be precisely located within the spatial domain. This basically means that the source of evidence is a point source — and the measurement operator is some kind of Dirac impulse at a specific location in time and space. Then, an 'area of influence' is either assumed based on model correlation lengths or it is computed with the help of heuristics (*e.g.* streamline simulation in case of subsurface

flows). From this information a compact, smooth 'damping operator' for the model space is constructed and used to remove covariance structure beyond the area of influence.

However, this approach does not consider measurements which have a smoothing or averaging nature. Examples are seismic measurements, but also dynamic data for diffusive systems in general. Due to such smoothing properties, small scale features (compared to the support of the measurement operator) may not be resolved by the evidence, and therefore cannot be reliably identified. However, common covariance localisation approaches do not take this into account. The effects of a diffusive measurement are demonstrated in the following example, and a way to mitigate them is discussed.

Consider a signal $s(x)$ consisting of two superposed sine waves over the domain $A = [0, 1]$, phase shifted by $a_{1,2} \in [0, 1]$ and amplified by $b_{1,2} \in [0, 1]$. It is described by the function

$$\begin{aligned} s(x) &= f(x, a_1, a_2, b_1, b_2) \\ &= \sin((10 + 3b_1)(x - a_1)) + (0.3 + 0.2b_2)\sin(200(x - a_2)). \end{aligned} \tag{5.1}$$

The first sine wave is of large scale (longer wavelength, larger amplitude), while the second one has a smaller scale. Two examples of such a signal are given in Fig. 5.1. Everywhere in this example a simple discretisation of A consisting of 100 equidistant points is used. Assume now that the actual signal is unknown, and one plans to identify it from evidence obtained by one of two possible measurement operations:

1. a Dirac measurement operator

$$D(s, x) = s\,\delta(x - 0.5), \tag{5.2}$$

   located precisely at $x = 0.5$, and

2. a Gaussian measurement operator

$$G(s, x) = s\,\sigma\sqrt{2\pi}\,\exp(-(x - \mu)^2/(\sqrt{2}\sigma)^2) \tag{5.3}$$

   with $\mu = 0.5$ and $\sigma = 0.02$ (see Fig. 5.2 for a plot of the discretised operator).

Both operators should measure the constant signal $s(x) = 1$ exactly. Therefore, the discretised form of the Gaussian operator is re-normalised

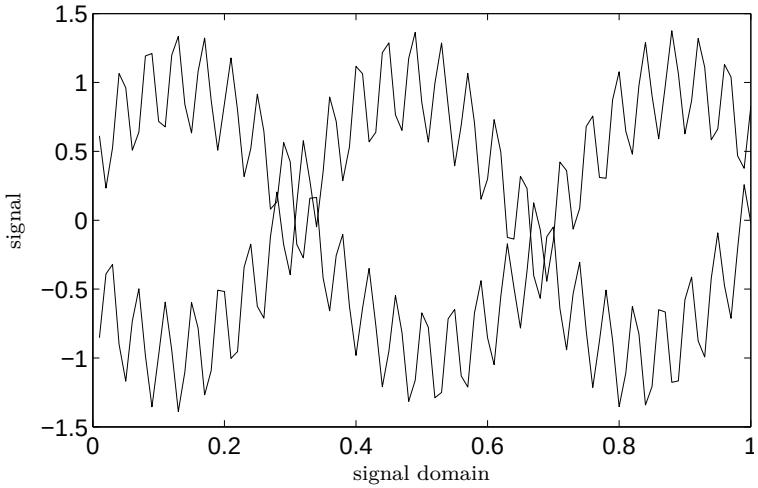

Figure 5.1.: Two example signals with the following coefficients: $f(x, 0.7, 0.2, 0.1, 0.4)$ and $f(x, 0.3, 0.1, 0.3, 0.5)$.

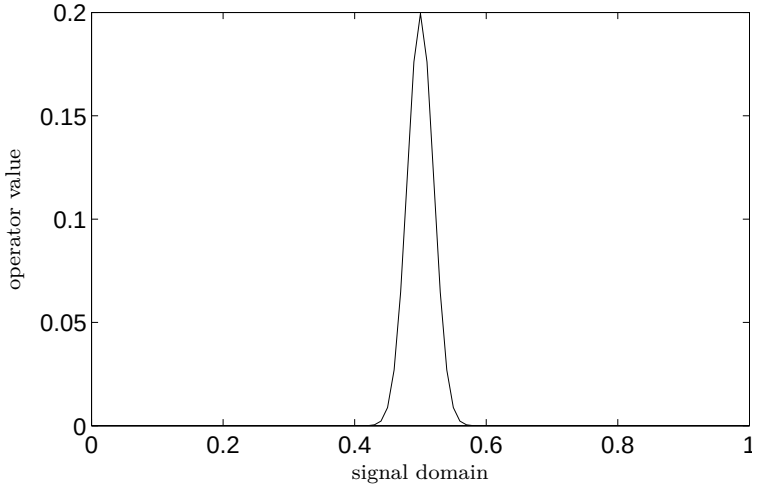

Figure 5.2.: Re-normalised Gaussian measurement operator $G(s, x)$ applied to the constant signal $s(x) = 1$. The maximum is located at $x = 0.5$ and more than 99.99% of the energy is concentrated in the interval $[0.4, 0.6]$.

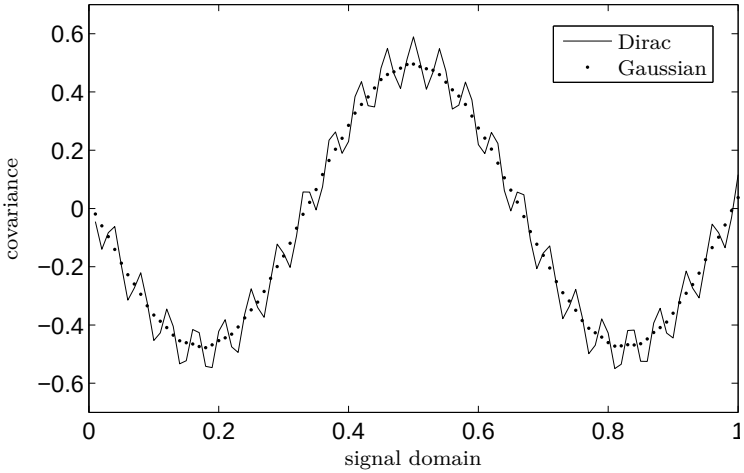

Figure 5.3.: Covariance for the two measurement operators, estimated from a large ensemble.

so that the sum of the coefficients is 1.0 — equal to the Dirac operator. For simplicity no measurement error is considered in this example.

An ensemble-based LCE method shall be used for signal identification. Therefore, the phase shifts and amplifications are considered as RVs with $a_{1,2}(\omega), b_{1,2}(\omega) \sim U(0,1)$. A sample of different phase shifts and amplifications is created as random realisations of these RVs. The realisations are inserted into Eq. (5.1) and evaluated over the domain, resulting a prior ensemble of discrete signals. With the prior ensemble we can estimate the covariance between the measurement result and the model variables (*cf.* Eq. (3.9)). Such an estimate is plotted in Fig. 5.3 for both measurement operators and for a large ensemble of $N = 10000$ realisations (therefore we can consider it as exact). There, one can see that the Dirac measurement carries information on the small signal scales, while the Gaussian one does not. While this may be somewhat expected, it is an important result: the smoothing operator simply cannot resolve small scale features of the signal, giving us no hope for identifying them with this setup. However — thinking of a real application — small scale features may also be of minor importance for the identification anyway, due to the smoothing nature of the system. Therefore, if the smoothing nature of the measurement operator or the system dynamics is known a priori, this

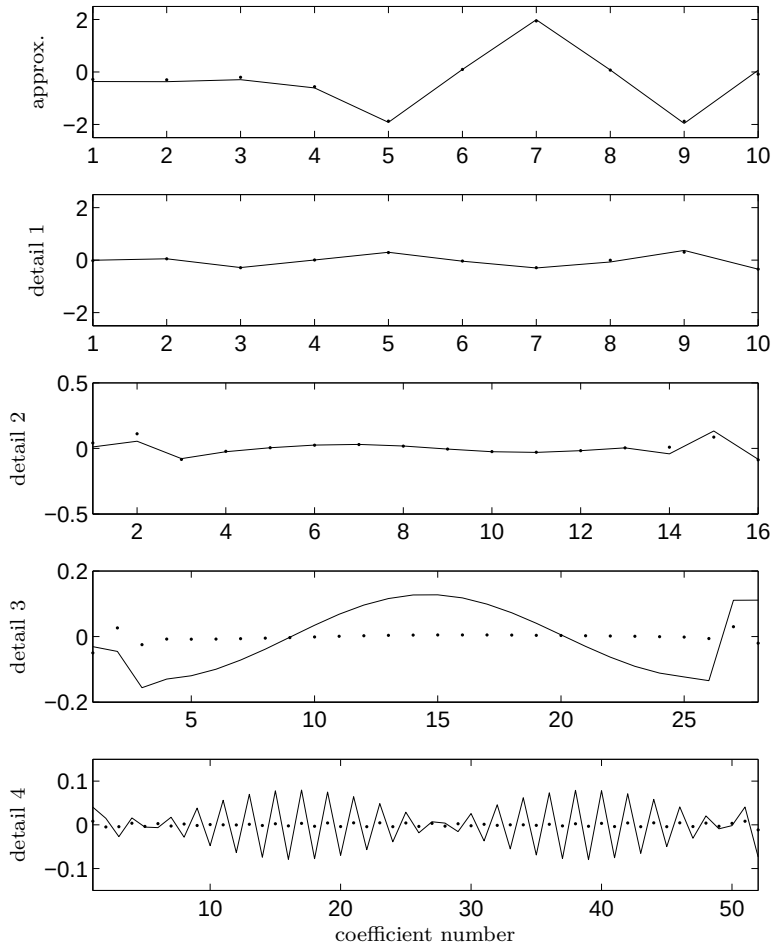

Figure 5.4.: Multiscale wavelet decomposition of the two estimated covariance functions from Fig. 5.3 using a Daubechies-3 wavelet. The dots represent the Gaussian operator covariance, whereas the lines correspond to the Dirac. The topmost plot represents coarsest structure, the following plots increasing details. Note the different numbers of coefficients for the levels.

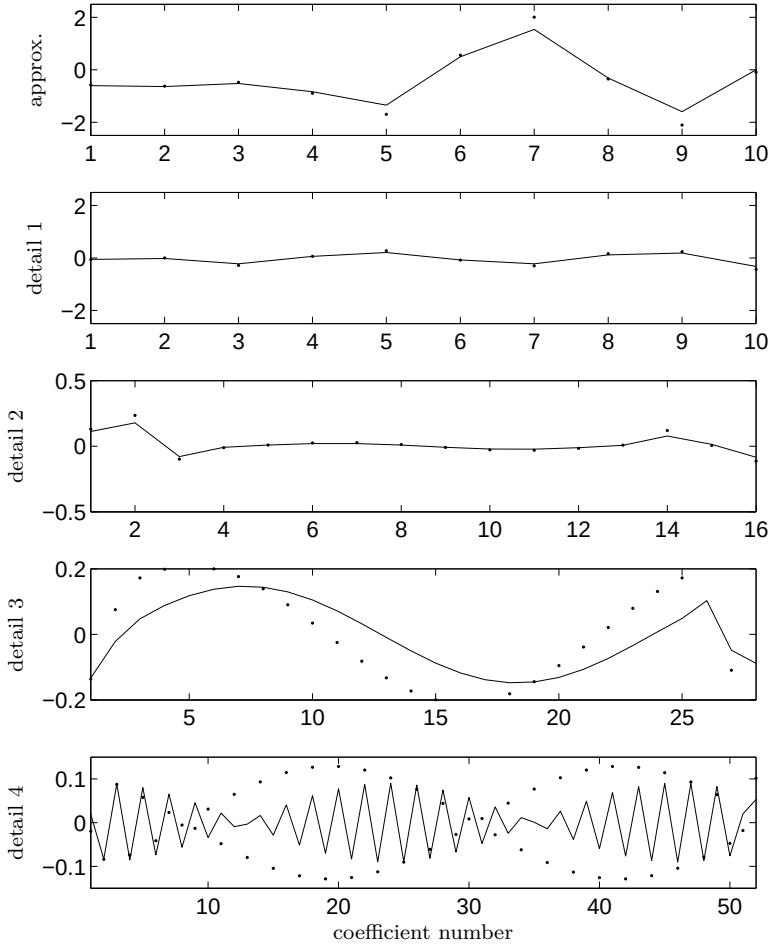

Figure 5.5.: Wavelet decomposition of the same covariance structure as Fig. 5.4, but estimated from a small ensemble of $N = 10$ realisations.

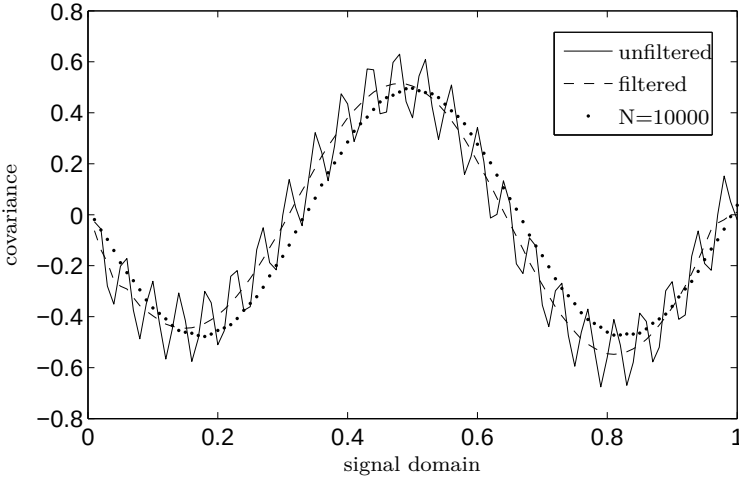

Figure 5.6.: Two reconstructions (unfiltered, filtered) of the covariance structure associated with the Gaussian measurement operator, estimated from a small ensemble of $N = 10$ realisations. Compared to the estimate obtained from a very large ensemble.

information should be included into the solution process for the inverse problem to further regularise it.

Multiscale wavelet analysis is a highly suitable tool to include this information into the process: consider Fig. 5.4, where a multiscale wavelet decomposition of the covariance functions of Fig. 5.3 is presented. The approximation coefficients and the detail coefficients of decomposition levels one and two are equal for both operators. They correspond to large scales in the covariance functions, meaning that both operators resolve those scales equally well. However, with increasing detail coefficient level, smaller scales are represented. There a major difference between the Gaussian operator and the Dirac operator can be seen: for the Gaussian operator the detail coefficients of levels three and four are practically zero, while the Dirac measurement coefficients are still significant. The example demonstrates that — even with a large ensemble and no measurement noise — smale scales in the model may not be resolved when a smoothing measurement operator or diffusive system dynamics are involved. This motivates us to neglect these scales in the identification — which is easily accomplished by setting irrelevant wavelet coefficients to zero. This is the underlying principle of the wavelet-based covariance localisation approach presented in this section.

Removing such irrelevant coefficients has two aspects: (1) it may save us from trying to identify model scales that cannot be identified from evidence anyway, and (2) it has the potential to avoid small-scale sampling noise resulting from the small ensemble sizes used in practical applications. Both aspects may be of interest for geoscience applications. While the first aspect is quite obvious, the second one needs some discussion: Fig. 5.5 shows the wavelet decomposition of the same covariance estimate as given in Fig. 5.4, but estimated from a three orders of magnitude smaller ensemble of $N = 10$ members. By direct comparison with Fig. 5.4, one can see that this small ensemble is already sufficient to estimate large scale covariance structure sufficiently well (up to detail coefficients of level two in this example). On the other hand, small scale coefficients contain significant errors — especially for the Gaussian measurement operator — due to the small ensemble size. Therefore setting those erroneous coefficients to zero may result in an improved covariance estimate. This is demonstrated in Fig. 5.6 for the given example. There, the unfiltered covariance estimate obtained from the small ensemble is plotted. It is compared to a naively filtered variant wheren the detail coefficients of levels three and four are completely set to zero before performing an inverse wavelet trans-

form. The obtained covariance estimate is indeed comparable to the one computed from a much larger ensemble.

Note that traditional covariance localisation does not help at all with this example: since the decorrelation distance is essentially infinity, it will remove valid covariance structure. On the other hand some regularisation approach is necessary to mitigate sampling errors, and the wavelet-based localisation approach can do this.

5.2. Other Regularisation Methods and the Wavelet-based Approach

From the literature review in section 3.4.1.4 the main ideas for regularisation of stochastic LCE methods can be sorted into the following groups, based on the type of additional knowledge they introduce:

Regularised Sampling encompasses all methods that do not employ knowledge of the *model*, but of the probabilistic nature of the involved *method* — they aim at mitigating sampling errors when the sampling is performed.

Regularisation of Variance encompasses all methods that aim at mitigating a specific effect of sampling, and especially of small Monte Carlo ensembles: the ensemble variance may prematurely collapse. Most of these methods do so by increasing the variance of the ensemble, either heuristically or adaptively.

Regularisation by Locality encompasses all methods which employ some measure of distance and position *within the model* to regularise the problem. Thus, these are the only methods which use knowledge from the *model* itself. For example local updating, covariance localisation, as well as local averaging employ distances and position of measurements within a spatial model domain $\mathcal{X}_s$ for the regularisation.

It should be noted that some of the adaptive 'regularised sampling' methods are sometimes termed as 'localisation' approaches in literature, be-

cause they tend to have similar effects on the estimation. However, this seems to be not conclusive and should be considered with care: just because sampling errors tend to be more dominant when distant correlations are to be estimated — and this effect can be detected and diminished by such methods — they do not use additional knowledge on the problem *based on locality*. And this is certainly the defining feature of any *localisation* approach.

The wavelet-space based regularisation method discussed in this section is certainly a covariance localisation approach, since it requires a notion of distance and is applied to covariance (or Kalman gain) estimates. By filtering out local small-scale noise it also has a strong connection to local averaging approaches. Additionally, due to the separation of scales, it can effectively filter out small scale long-range correlations which are usually solely due to noise. Therefore it could be seen as a combination of several regularisation approaches. Interestingly, this implies that it still could be worthwile to combine it with regularised sampling and inflation, since these aim at mitigating different sources of error, as discussed above.

5.3. Related Work

The approach presented by Chen et al. [89] is a close relative to the ideas developed in this section. However, there a screening approach is used to filter out 'unreliable' covariance estimates in wavelet space. It is mainly based on knowledge from the method domain[1] and therefore includes additional knowledge from the model domain only in a limited way.

5.4. Connected Publications and Discussion

The state of the work has been published by Pajonk et al. [5]. It can be seen as a mixed spatial-spectral regularisation technique for probabilistic linear Bayesian updating. However, the current results of the research did not justify an inclusion into this thesis. Especially the design of the

[1] The knowledge used is that sampling errors occur due to the probabilistic nature of the method.

wavetlet-space localisation operator proves to be crucial: an aggressive, instable construction may quickly lead to filter divergence, while a non-optimal construction leads to similar results already obtained by usual covariance localisation schemes in combination with inflation approaches.

Chapter 6.

An Ensemble Kalman Filter and Evolution Strategy Hybridisation

Over the course of this thesis, research has been conducted towards a conceptual combination of ensemble Kalman filtering and *evolution strategies* (ES, [314, 298, 315, 57, 58]) with application to inverse problems.

6.1. Motivation

The motivation to do this clearly comes from the practical application point of view: evolutionary optimisation methods have proven their usefulness in real day-to-day applications over some time (*e.g.* [313]), which is something the EnKF still has to achieve. On the other hand, the EnKF has shown remarkable performance in some large-scale applications (*e.g.* [159, 59, 118, 86, 87] [60, chapter 12]).

Both ES and the EnKF have in common that they represent their solution of the problem by a set of possible solutions rather than a single guess. This approach is called population-based or ensemble-based. In the EnKF the set of realisations has a statistical meaning: they are a discretisation of a PDF for the state; the population in ES generally does not. This is of course a drawback when it comes to uncertainty quantification: in contrast to the EnKF we cannot directly compute any statistical moments from an ES population as it is not designed to represent a pdf; we can only obtain a best-guess estimate of the objective values.

Integrating an ensemble forward in time and assimilating data piecewise is a concept that is widely unknown to ES due to its origin in numerical optimisation: there we do not generally have a time dimension in the problem that can be exploited by the algorithm. An individual is scored by computing an objective function; if this involves integrating the system over a time period and computing measurements does not influence the algorithm. Therefore these algorithms cannot operate on the state of a model, only on its parameters. The ES have a virtual time dimension related to iterations of the optimisation process. They are called generations, again in analogy to the biological concept.

In contrast to ES, the EnKF has an intrinsic time dimension; it changes the model states of the ensemble members and their parameters at every assimilation step and computes the members forward in time until the next assimilation step can take place. The update is performed using a linear data mismatch term and the combined covariance structure of the ensemble and the measurements. These are used to create an update term for each ensemble member at every assimilation step. Thus we can regard the Kalman gain matrix, in comparison to ES, as a combined scoring and recombination operator.

The purpose of the mutation operator of ES can be compared to the measurement ensemble created for the analysis step in the EnKF: Random changes are introduced to measurements in the EnKF that are similar to mutations in ES. However, in the EnKF mutations are not performed on the ensemble members themselves but on the input data.

6.2. Evolution Strategies

Developed in the engineering community, evolution strategies are heuristical optimisation algorithms. The research in ES was driven by complex, practical optimisation problems on which classical algorithms failed. They are part of the bigger class of *evolutionary algorithms* (EAs, [80]), a sub-field of bionics (*e.g.* [352]). Thus they are a synthesis of biology and technology; other examples in computer science are *artificial neural networks* [304] and *swarm intelligence* approaches [204, 37, 38]. Closely related to ES are the *genetic algorithms* (GAs, [140, 84, 85]).

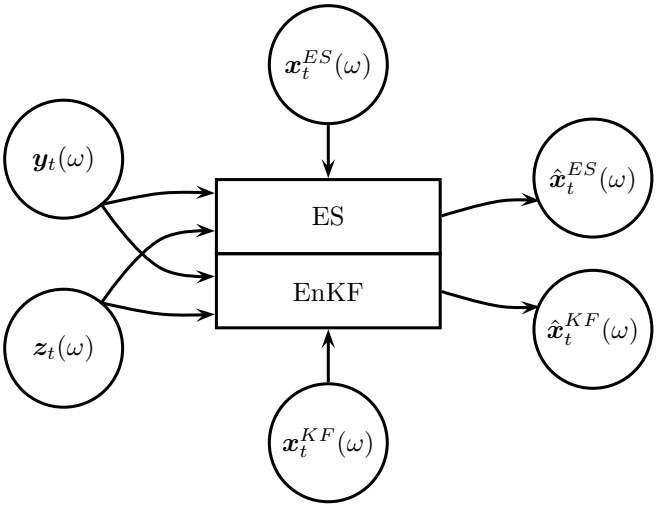

Figure 6.1.: A parallel hybrid EnKF/ES scheme. This diagram shows the basic idea of one step of a *parallel* hybrid EnKF/ES scheme. Both algorithms use the same data but operate on different subsets of $\boldsymbol{x}_t(\omega)$.

The central idea of evolutionary optimization algorithms is that they do not create and track a single solution to an optimisation problem by improving a first guess iteratively. Their approach is to use an entire *population* of different solutions which are called *individuals*. These solutions are represented by their genome, which is a vector of values describing a solution. During the optimisation process this whole population is improved iteratively using the main principles of evolution: *relative fitness*, *selection*, *mutation* and *recombination*. For further details, the reader is referred to above publications.

6.2.1. The Covariance Matrix Adaption Evolution Strategy for Robust Optimisation

The *covariance matrix adaption evolution strategy* (CMA-ES, [156, 154, 155]) is a modern, robust and quasi-parameter-free algorithm highly suit-

able for the hybridisation approach. For example, Heidrich-Meisner and Igel [161] show that this algorithm is quite robust w.r.t. noise, choice of hyper-parameters and initial value. Detailed results on noisy test problems are given by Auger et al. [35], and many other successful application cases exist [153].

The basic idea of CMA-ES is to adapt the mean and covariance matrix of a multi-variate normal distribution locally to a problem by successively improving it with certain low-rank updates. This is somewhat a resemblance of the updates of the inverse Hessian matrix in quasi-Newton methods like BFGS [100]. In fact, Hansen [154, p. 78] argues '[...] that the optimal covariance matrix equals the inverse Hessian matrix, up to a constant factor. Consequently, the adaptation mechanism should aim to approximate the inverse Hessian matrix.' The new mean and the covariance updates are computed in a certain way from the current population members (like in the EnKF, so to say) as well as a history of previous population members. A new ES population is then generated by sampling from the adapted multivariate normal distribution.[1]

The EnKF and the CMA-ES can be compared more specifically. They both try to estimate a covariance matrix from the population/ensemble. However, the approaches are quite different: the EnKF does not use information from historical ensembles, whereas the CMA-ES does so to stabilise the estimation of the covariance. The EnKF creates new population/ensemble members by a weighted linear combination of the previous ones — with random fluctuations due to the random measurement ensemble — whereas the CMA-ES creates completely new population/ensemble members by randomly sampling them from a distribution which has been estimated from the historical evolution of the population.

There are basically two possible approaches for hybridisation: the ES and EnKF can be run in parallel (on different objective values, of course) or serially. It is worthwile to allow for both approaches for maximum flexibility. While the serial approach is conceptually straightforward — simply apply first one method, e.g. the ES, to one set of objective values and afterwards the other method to a different set — the parallel approach needs some explanation. The basic idea is sketched in Fig. 6.1: split the model $\boldsymbol{x}_t(\omega)$ into two disjunct sets, $\boldsymbol{x}_t^{ES}(\omega)$ and $\boldsymbol{x}_t^{KF}(\omega)$. The choice of

[1] For details see the work of Hansen [154].

splitting is up to the user and should be performed w.r.t. the strength of either algorithm. Then, at each time new evidence $\boldsymbol{z}_t(\omega)$ is available, each of the algorithms receives this — as well as the simulated data $\boldsymbol{y}_t(\omega)$ — and can compute an update to its set. However, analogous to [167], one may treat $\eta(\omega)$ or $\xi(\omega)$ by one of the methods, *e.g.* the ES, to improve the performance of the other one treating $\boldsymbol{x}_t(\omega)$. In ES this is a common approach: parameters of the algorithm, for example mutation strength or probability, are treated as part of the genome of an individual. They are evolved using the same principles that modify the solution to the problem itself. This is the basis for the so-called *self-adaptivity* of evolutionary algorithms: they can adopt their parameters to the problem. This bears some resemblance of hierarchical Bayesian [135, chapter 5] approaches, with an example being the HEnKF shortly mentioned in section 3.4.1.3.

It is clear that to make such a hybrid approach robust, each of the involved algorithms has to be able to cope with what is called a *noisy* objective (*e.g.* [327, section 2.3]). If there are mutual interdependencies between variables in the two sets these clearly will have effects on the algorithms when operated this way.

6.3. Related Work

A related comparison on evolutionary algorithms and filters, which also resulted in a hybrid algorithm, is given in [165]. There, the authors compare GA and probabilistic filters and develop a hybrid strategy [166]. In [167] they discuss this method, which aims at estimating unknown parameters of a measurement and model error distribution using a genetic algorithm. The results are subsequently employed in a probabilistic filter algorithm. They call this a 'self-organising state space model'.

Approaches related to the presented one are the combination of ideas from Evolutionary Algortihms and Markov Chain Monte Carlo [109], and, more generally, of evolutionary and statistical ideas in, for example, [40].

6.4. Connected Publications and Discussion

The state of the work and applications of it have been published by Pajonk et al. [2, 3], Schulze-Riegert et al. [12, 13], and Pajonk et al. [4]. The conceptual connection of EnKF and ES has lead to a robust, flexible software framework which can handle a wide amount of optimisation tasks — especially in the context of hydrocarbon reservoir optimisation.

Detailed research results have been excluded from this work since the concept does not directly fit the main theme of 'linear spectral methods', and therefore would cause unnecessary disturbance for the reader. However, the reader is encouraged to review above mentioned publications for further information.

Chapter 7.

Conclusions

In this work a Bayesian framework for inference has been presented which is based on random variables as the basic building blocks. This approach allows for a natural, direct treatment of uncertainties in the context of identification and, by direct extension, optimal control problems. The major existing lines of methodological research have been highlighted and connected to this framework. This review turned up some prospective areas of research as well as some more apparent ones.

For the area of linear Bayesian identification a new approach has been presented which allows for direct, sequential inference solely based on stochastic spectral expansions. The approach has been compared to existing, related ones on several numerical examples. On a scalar example the relation to probabilistic relatives and to fully non-linear Bayesian updating has been discussed. On Gauss-linear problems it reproduces the Kalman filter results, as theory predicts. Two related, approximative approaches presented by other authors have been included into the comparison. On a more complex example the new approach has advantages over probabilistic methods when estimating variance, but especially also higher moments and full probability densities. This is mainly due to the usage of orthogonal spectral expansions, which have — depending on the application — favourable convergence properties. With this new, fully spectral approach no detour to sampling methods is necessary to infer the posterior spectral expansion — also in the context of *sequential* identification of dynamical systems. Additionally, real time applications may benefit from it, because the implementation of this approach is fully deterministic.

The limitations of this approach, as well as of two related ones, have been investigated on parameter estimation tasks. It has been demonstrated that a non-linear connection between variables may cause problems for these methods if the involved uncertainties are too high — depending on the specific application.

Additionally, a conceptual combination of ensemble-based sequential identification methods and population-based optimisation approaches has been developed over the course of the thesis. While the results are quite satisfying and have led to improvements in complex, practical inference workflows the approach is not strongly connected to the main theme of this thesis and therefore only briefly presented. Readers specifically interested in this topic are kindly referred to the publications referenced in the main text.

7.1. Outlook

As usual in research, only some possibilities for the presented inference approach could be investigated. Therefore we have picked out the — in our opinion — most immediate directions of research and have tried to properly investigate them. Clearly a lot of work remains to be done — some of which is given as an outlook here.

The SRPCU approach itself does not have a strong dimensional dependence. Similarly to ensemble schemes, local updating approaches may be developed to support large observation sets. However, the growth of the underlying spectral expansion may limit the applicability. It will therefore be crucial to combine this inference approach with adaptive subspace selection schemes for the stochastic representation. In that context, also the 'correct' LPCU implementation may become competitive for sequential applications and needs to be re-considered.

With the PCE being just one example of a suitable series expansion, the combination with gPCE approaches may be important for certain applications. Also the combination with stochastic collocation approaches for the forward solution may be worthwhile (*cf.* [369]). As to extensions towards non-linear identification, iterative variants of this scheme are promising and do not entail major conceptual effort or modifications

(*cf.* [340, Eq. (3.51)]). This should further enhance the applicability to non-linear problems. Also the investigation of regularisation techniques similar to the ones already popular with probabilistic implementations like the EnKF may further improve the performance of this method. Lastly, the investigation of pre-multiplication square root schemes (in contrast to the presented one) may result in advantageous methods.

In addition to the spectral identification method, the concept of a specific multi-scale covariance localisation approach for probabilistic sampling LCE has been shortly presented. It is based on the usual assumptions of relative spectral importance of large-scale, slowly varying components of a signal. Using a multi-scale wavelet decomposition, such parts of a signal could be discriminated and specially designed localisation functions could be applied to different parts of a signal. However, this line of research has not yet led to results justifying an extensive discussion, which is why it demands only a small share of this thesis. The specific difficulty with the approach is the construction of the localisation operator, where further research and numerical tests are necessary.

Appendix A.

Random Variables

This appendix summarises some important properties and results for RVs which are used throughout the text.

A.1. Random Fields and Random Processes

It is usual to denote an RV having a spatial extent as a *random field* (RF) and an RV having a time extent as a *random process* (RP). Obviously, combinations are possible — which will be denoted in general as random fields. To add to the confusion, some authors also use the notion of random process for random fields.

Since the model and data spaces introduced in Eq. (2.1) on page 7 are assumed to be Banach spaces (and not simply the d-dimensional real or complex numbers), there are several *non-equivalent* ways to define generalised RFs and RPs (*cf.* [318, 319, 168]). This work follows the approach of Holden et al. [168, esp. Eq. (2.2.21)], where random processes generalised w.r.t. the ω-argument and pointwise defined on the remaining parameters are introduced. This fits the necessities of this work, as $\boldsymbol{x}_s$ is — as discussed in section 2.1.4 on page 13 — usually a subset of $\mathbb{R}^d$, and the RFs therefore do not need to be generalised w.r.t. this parameter (but see [318, 319] for an approach in this direction). However, in the construction of Holden et al. [168] used here, the model and data spaces need to be Hilbert spaces — but this poses no practical restrictions on us.

RFs/RPs are are written as random spatial functions. This is expressed

by writing them in boldface (remember that $\mathcal{X}$ may already be a function space;*cf.* section 2.1.4):

$$\boldsymbol{r}(\omega) \in L_0(\Omega; \mathcal{X}). \tag{A.1}$$

Regarding Eq. (2.10) on page 11, for RPs (which take values in $\mathcal{X}_q$) we usually write $r_t(\omega)$ and for time dependent RFs $\boldsymbol{r}_t(\omega)$. This setup allows us to distinguish between RVs, RFs and RPs only by notation. Note that a similar construction, of course, applies to the data domain $\mathcal{Y}$, as well as to $\mathcal{M}$.

A.2. The Connection of Random Variables and Information Content

The connection between RVs and their *information content* is the tool for representing 'incomplete information caused by errors' in our stochastic approach to modelling uncertainty. This connection is established via the so-called σ-algebra *generated* by an RV $r(\omega)$:

$$\sigma(r) := \{B \in \mathfrak{S} \,|\, B = r^{-1}(E) \text{ where } E \in \mathfrak{B}(\mathcal{X})\}. \tag{A.2}$$

Obviously, $r(\omega)$ is $\sigma(r)$-measurable, but it is also $\mathcal{A}$-measurable for any σ-algebra $\mathcal{A}$ with $\sigma(r) \subset \mathcal{A}$ (see also [68, p. 4]).

The generated σ-algebra can be coarse or fine, meaning that it contains less or more different events for which a probability is known. It is in this sense that a σ-algebra can be seen as to represent the *information content* of an RV. For example, Billingsley [62, p. 57f] discusses the significance of sub-σ-algebras (sub-σ-fields in his denomination) in probability theory, and how they can be seen as to represent partial information. Applebaum [27] aims to combine information and probability theory, at least on discrete sample spaces. Gray [146, p. 173ff] discusses the concept of *relative entropy* or *divergence* of two probability measures on the same probability space. Kullback and Leibler [218] is an early reference on this topic, where the *Kullback-Leibler divergence* between two probability measures is introduced as a measure of 'information gain'. Gourieroux and Monfort [144, section 3.3] describe a popular absolute measure of information, the *Fisher information*, and how it is related to the Kullback-Leibler divergence. Ben-Naim [46, p. xvii] even comes to the following conclusion:

> Finally, I believe that the time has come to reach the inevitable conclusion that "entropy" is a misnomer and should be replaced by either missing information or uncertainty. These are more appropriate terms for what is now referred to as "entropy."

Obviously, different RVs create different push-forward probability measures and σ-algebras. Our aim is to use RVs to describe a probability measure which represents our state of knowledge on the space $\mathcal{M}$. Such knowledge may come from objective data, *e.g.* seismic measurements of some underground structure, information on accuracies of involved measurement devices, statistical information obtained from problems somehow 'similar' to the one at hand — but also subjective 'engineering knowledge' coming from several years of work experience in practical applications. This knowledge may be quite complex, and it is often not easy at all to formally capture it into RVs. Nonetheless, this step is a very important one in the whole process, especially when data is sparse. Stuart [335, p. 462] clearly demonstrates this.

A.3. The Connection of Random Variables and Probability Densities

Section 2.3.1 shows that an RV $r(\omega)$ creates a push-forward probability measure $r_*\mathbb{P}$ on $\mathcal{X}$. Given a reference measure μ on $\mathcal{X}$, with $r_*\mathbb{P}$ being *absolutely continuous* w.r.t. μ (usually written as $r_*\mathbb{P} \ll \mu$; *e.g.* [112, p. 279] or [68, p. 9]), by the *Radon-Nikodým-theorem* (*e.g.* [33, p. 65]) the measure $r_*\mathbb{P}$ possesses a *probability density function* (PDF) $p(\cdot)$ w.r.t. μ in the sense that

$$\forall E \in \mathfrak{B}(\mathcal{X}) \,:\, \int_E r_*\mathbb{P}(\mathrm{d}x) = \int_E p(x)\mu(\mathrm{d}x). \tag{A.3}$$

This is usually written as

$$\frac{\mathrm{d}r_*\mathbb{P}}{\mathrm{d}\mu}(x) = p(x), \tag{A.4}$$

and is called the *Radon-Nikodým-derivative*. If μ happens to be the Lebesgue measure, then $p(x)$ is simply called *the* probability density function

[68, p. 9] — although it can be a *generalised* function in the sense of distributions [208].

As already mentioned the spaces $\mathcal{X}$ and $\mathcal{Y}$ are assumed to be Banach spaces. For this setting the reference probability measure μ cannot be the Lebesgue measure, due to the possibly infinite dimensions the Banach spaces may have: The infinite-dimensional Lebesgue measure, which we would need to be able to find 'the' probability density functions for $r_*\mathbb{P}$ does not exist. However, many measures have a density with respect to a Gaussian measure (*cf.* [335, p. 527]), and μ is therefore often assumed to be such a measure[1]. In addition, Tarantola [340, p. 11] hints that even an infinite 'volume' of $\mathcal{X}$ and $\mathcal{Y}$ — resulting in an infinite reference measure μ, which is therefore no longer a probability measure — does not generally cause problems. However, since the primary tool of this work are RVs this aspect is not discussed any further.

PDFs are the bridge to classical probability theory, which is strongly founded on mathematical measure theory and therefore employs PDFs and probability measures as its primary objects. In contrast, in this work the RVs themselves are the primary objects of interest.

A.4. An Integration Algebra of Random Variables

In addition to being a vector space, RVs $r(\omega) \in L_\infty(\Omega;\mathbb{R})$ form an associative and commutative C^*-algebra, often denoted as $\mathcal{B}$ (*cf.* [317, chapters 8 and 14] or [258, section 13.4][2]). This algebra is itself a Banach space, possesses an algebra-anti-automorphism called *involution* (usually denoted by a star, $(\cdot)^*$, and the algebra multiplication is continuous. One usually assumes that the algebra has a distinguished element

$$e \in \mathcal{B} \tag{A.5}$$

[1] A Borel measure μ on a separable Banach space $\mathcal{V}$ is called Gaussian if the push-forward of μ by any non-zero linear functional in the continuous dual space to $\mathcal{V}$ is a Gaussian measure on $\mathbb{R}$. The reader interested in details on this is kindly referred to [94, chapter 1].

[2] In this work we exclusively deal with real valued RVs.

called the *unit* element. In this algebra we have defined a *multiplication* operation

$$\forall a, b \in \mathcal{B} : \; a \cdot b \mapsto c \in \mathcal{B}, \tag{A.6}$$

positive elements $y \in \mathcal{B}$ with

$$y = x^* x, x \in \mathcal{B}, \tag{A.7}$$

(usually written as $y \geq 0$) and also k-th roots (*cf.* [258, p. 65]), thereby introducing the additional structure which is necessary to allow for RVs as primary objects of interest in probability theory.

Additionally, one assumes that a certain positive linear functional (an abstract integral)

$$\mathbb{E}(\cdot) \in \mathscr{L}(\mathcal{B}; \mathbb{R}) \tag{A.8}$$

called the *expectation* is defined. Concretely, the expectation of any measurable function f of an RV is defined as

$$\mathbb{E}\left(f(r)\right) := \int_{\Omega} f(r(\omega)) \, \mathbb{P}(\mathrm{d}\omega). \tag{A.9}$$

For brevity any extensive discussion about the properties of this integral and the connection to measures $\mathbb{P}$ is omitted here. The reader is referred to the respective literature (*e.g.* [33]). Let us from now on assume that all technical requirements for well-posedness of the involved functions and measures are fulfilled. The expectation functional is used to define an inner product (*cf.* Eq. (A.12)), and $\mathcal{B}$ may be completed w.r.t. this inner product [317]. Note that for $\mathcal{B} = L_\infty(\Omega; \mathbb{R})$ this yields $L_2(\Omega; \mathbb{R})$. This important special case is discussed further in appendix A.5.

For RFs $\boldsymbol{r}(\omega) \in L_0(\Omega; \mathcal{X})$, above definitions can be extended in the following way: assume that $\mathcal{X}_q \subseteq \mathbb{R}^n$ (until now it has been assumed that it is a generic vector space; however, this is practically no restriction for real applications), and for each n consider $L_0(\Omega; \mathbb{R}) \otimes \mathcal{X}_s$. The expectation defined on this product for $\psi_i \in L_0(\Omega; \mathbb{R})$ and $\boldsymbol{x}_s \in \mathcal{X}_s$ becomes (also for each n)

$$\mathbb{E}\left(\sum_i \psi_i \otimes \boldsymbol{x}_s\right) := \sum_i \mathbb{E}\left(\psi_i\right) \otimes \boldsymbol{x}_s. \tag{A.10}$$

Here $\otimes$ denotes as usual the *tensor product* or *outer product*. For a more complete description see also the book of Janson [184, appendices B and C].

A different approach to define an algebra of RVs is presented in the book of Springer [328]. There, distribution functions for algebraic functions of RVs are derived via integral transforms like Fourier, Laplace and especially Mellin. RVs are — again — not treated as primary objects of interest, though.

A.5. The Hilbert Space of Random Variables

Under certain circumstances it is useful to consider the RVs $r(\omega)$ living in the Lebesgue space $L_2(\Omega;\mathcal{M})$, which is a subspace of $L_0(\Omega;\mathcal{M})$ defined previously. This is the space of RVs with finite second moment or — equivalently — of finite variance (*cf.* Eq. (A.20))

$$\mathrm{tr}(\boldsymbol{C}_r) < \infty. \tag{A.11}$$

Given that $\mathcal{M}$ is a Hilbert vector space with inner product $\langle\cdot|\cdot\rangle_{\mathcal{M}}$ these RVs form a complete Hilbert space of measurable functions (*e.g.* [168, Eq. (2.2.21)]). The expectation operation defined in Eq. (A.8) defines an inner product on $L_2(\Omega;\mathcal{M})$ via

$$\langle r_1|r_2\rangle_{L_2(\Omega;\mathcal{M})} := \mathbb{E}(\langle r_1|r_2\rangle_{\mathcal{M}}), \tag{A.12}$$

and the norm induced by this inner product is

$$\|r\|_{L_2(\Omega;\mathcal{M})} := \sqrt{\langle r, r\rangle}_{L_2(\Omega;\mathcal{M})}. \tag{A.13}$$

Informally speaking this inner product can be used to measure angles (representing correlation) between RVs, which will become highly important later on.

A similar L_2-setting is described in Luenberger [245, p. 79ff], but without any connection to probability measures on a joint model-data space $\mathcal{M}$ and no discussion on the impact of a continuous spatial extent of the model.

A.6. Elementary Functions of Random Variables

For the sake of completeness and to introduce notation some of the most important examples of measurable functions are given in this section. Most of these functions have a form similar to

$$\mathbb{E}\left(r^n\right) := \int_\Omega r(\omega)^n \, \mathbb{P}(\mathrm{d}\omega), \; n \in \mathbb{N}, \tag{A.14}$$

which is called the nth *moment* of $r(\omega)$. The special case of the first moment,

$$\bar{r} := \mathbb{E}\left(r\right) = \int_\Omega r(\omega) \, \mathbb{P}(\mathrm{d}\omega), \tag{A.15}$$

is called the *expected value* or *mean* of the RV $r(\omega)$. Note that this quantity can be considered as a constant, deterministic offset in stochastic space. Therefore it is sometimes convenient to consider the 'random part' of an RV alone, so we define (in case $\bar{r}(\omega) < \infty$)

$$\tilde{r}(\omega) := r(\omega) - \bar{r}. \tag{A.16}$$

For RVs $\tilde{r}(\omega)$, the mean is obviously zero, thus we call them *centred*. This allows us to define the nth *centred moment* of $r(\omega)$:

$$\mathbb{E}\left(\tilde{r}^n\right) := \int_\Omega \tilde{r}(\omega)^n \, \mathbb{P}(\mathrm{d}\omega), \; n \in \mathbb{N}. \tag{A.17}$$

The second centred moment $\mathbb{E}\left(\tilde{r}^2\right)$ is also known as the *variance* of $r(\omega)$.

It has been already mentioned that, due to the spatial and/or time extent of the system under consideration, one often works with random fields. Above definitions extend quite naturally to this more general case. In the following we limit the notation to spatial dependence as temporal dependence is technically the same. Let us define for $\boldsymbol{r}_1(\omega), ..., \boldsymbol{r}_n(\omega) \in L_2(\Omega; \mathcal{M}), n \in \mathbb{N}$ the nth *centred moment* by

$$\boldsymbol{M}^n_{\boldsymbol{r}_1 \ldots \boldsymbol{r}_n} := \mathbb{E}\left(\otimes_{j=1}^n \tilde{\boldsymbol{r}}_j\right), \tag{A.18}$$

a tensor of order n. As a special case of this, for two RVs $\boldsymbol{r}_1(\omega), \boldsymbol{r}_2(\omega) \in L_2(\Omega; \mathcal{M})$, we have

$$\boldsymbol{C}_{\boldsymbol{r}_1 \boldsymbol{r}_2} := \boldsymbol{M}^2_{\boldsymbol{r}_1 \boldsymbol{r}_2} = \mathbb{E}\left(\tilde{\boldsymbol{r}}_1 \otimes \tilde{\boldsymbol{r}}_2\right), \tag{A.19}$$

which we call the *cross-covariance* of the RFs $\boldsymbol{r}_1(\omega)$ and $\boldsymbol{r}_2(\omega)$. The special case of $\boldsymbol{r}_1(\omega) = \boldsymbol{r}_2(\omega) = \boldsymbol{r}(\omega)$ given by

$$\boldsymbol{C}_{\boldsymbol{r}} := \boldsymbol{C}_{\boldsymbol{r}\boldsymbol{r}} \tag{A.20}$$

is called *covariance* of $\boldsymbol{r}(\omega)$.

Appendix B.

Polynomial Chaos Expansion

This appendix lists additional material which is helpful to understand and implement the PCE in numerical schemes. See also, *e.g.* , the introductory part of Le Maître and Knio [224], Matthies [256], or Xiu [363] for related information.

B.1. Multi-Indices

In the PCE formulation, the need for multi-indices of arbitrary length arises. Formally they may be defined by

$$\alpha = (\alpha_1, \ldots, \alpha_j, \ldots) \in \mathcal{J} := \mathbb{N}_0^{(\mathbb{N})}, \tag{B.1}$$

which are sequences of non-negative integers, only finitely many of which are non-zero. As by definition $0! := 1$, the following expressions are well defined:

$$\begin{aligned} |\alpha| &:= \sum_{j=1}^{\infty} \alpha_j, \\ \alpha! &:= \prod_{j=1}^{\infty} \alpha_j!, \\ \ell(\alpha) &:= \max\{j \in \mathbb{N} \,|\, \alpha_j > 0\}. \end{aligned} \tag{B.2}$$

B.2. Hermite Polynomials

As there are different ways to define — and to normalise — the Hermite polynomials, a specific way has to be chosen. In applications with probability theory it seems most advantageous to use the following definition [164, 168, 184, 247]:

$$h_k(t) := (-1)^k e^{t^2/2} \left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^k e^{-t^2/2}; \; \forall t \in \mathbb{R}, \, k \in \mathbb{N}_0, \tag{B.3}$$

where the coefficient of the highest power of t — which is t^k for h_k — is equal to unity.

The first five polynomials are

$$\begin{aligned} h_0(t) &= 1, \qquad h_1(t) = t, \qquad h_2(t) = t^2 - 1, \\ h_3(t) &= t^3 - 3t, \qquad\qquad h_4(t) = t^4 - 6t^2 + 3, \end{aligned}$$

and the recursion relation for these polynomials is

$$h_{k+1}(t) = t\, h_k(t) - k\, h_{k-1}(t); \quad k \in \mathbb{N}. \tag{B.4}$$

These are orthogonal polynomials w.r.t. the standard Gaussian probability measure Γ, where $\Gamma(\mathrm{d}t) = (2\pi)^{-1/2} e^{-t^2/2}\, \mathrm{d}t$ — the set $\{h_k(t)/\sqrt{k!} \,|\, k \in \mathbb{N}_0\}$ forms a complete orthonormal system (CONS) in $L_2(\mathbb{R}, \Gamma)$ — as the Hermite polynomials satisfy

$$\int_{-\infty}^{\infty} h_m(t)\, h_n(t)\, \Gamma(\mathrm{d}t) = n!\, \delta_{nm}. \tag{B.5}$$

Multi-variate Hermite polynomials will be defined right away for an infinite number of variables, i.e. for $\boldsymbol{t} = (t_1, t_2, \ldots, t_j, \ldots) \in \mathbb{R}^{\mathbb{N}}$, the space of all sequences. This uses the multi-indices defined in appendix B.1: For $\alpha = (\alpha_1, \ldots, \alpha_j, \ldots) \in \mathcal{J}$ remember that except for a finite number all other α_j are zero; hence in the definition of the multi-variate Hermite polynomial

$$H_\alpha(\boldsymbol{t}) := \prod_{j=1}^{\infty} h_{\alpha_j}(t_j); \quad \forall \boldsymbol{t} \in \mathbb{R}^{\mathbb{N}}, \, \alpha \in \mathcal{J}, \tag{B.6}$$

except for finitely many factors all others are h_0, which equals unity, and the infinite product is really a finite one and well defined.

The space $\mathbb{R}^{\mathbb{N}}$ can be equipped with a Gaussian (product) measure [164, 168, 184, 247], again denoted by Γ. Then the set $\{H_\alpha(\boldsymbol{t})/\sqrt{\alpha!} \mid \alpha \in \mathcal{J}\}$ is a CONS in $L_2(\mathbb{R}^{\mathbb{N}}, \Gamma)$ as the multivariate Hermite polynomials satisfy

$$\int_{\mathbb{R}^{\mathbb{N}}} H_\alpha(\boldsymbol{t})\, H_\beta(\boldsymbol{t})\, \Gamma(\mathrm{d}\boldsymbol{t}) = \alpha!\, \delta_{\alpha\beta}, \tag{B.7}$$

where the Kronecker symbol is extended to $\delta_{\alpha\beta} = 1$ in case $\alpha = \beta$ and zero otherwise.

B.3. The Hermite Algebra

Consider first the usual univariate Hermite polynomials $\{h_k\}$ as defined in Eq. (B.3). As the univariate Hermite polynomials are a linear basis for the polynomial algebra, i.e. every polynomial can be written as linear combination of Hermite polynomials, this is also the case for the product of two Hermite polynomials $h_k h_\ell$, which is clearly also a polynomial:

$$h_k(t) h_\ell(t) = \sum_{n=|k-\ell|}^{k+\ell} c_{k\ell}^{(n)} h_n(t) \tag{B.8}$$

The coefficients are only non-zero for integer $g = (k + \ell + n)/2 \in \mathbb{N}$ and if $g \geq k \wedge g \geq \ell \wedge g \geq n$ [247]. They can be explicitly given

$$c_{k\ell}^{(n)} = \frac{k!\, \ell!}{(g-k)!\, (g-\ell)!\, (g-n)!}\,, \tag{B.9}$$

and are called the structure constants of the univariate Hermite algebra.

For the multivariate Hermite algebra, analogous statements hold [247]:

$$H_\alpha(\boldsymbol{t}) H_\beta(\boldsymbol{t}) = \sum_\gamma c_{\alpha\beta}^\gamma H_\gamma(\boldsymbol{t}). \tag{B.10}$$

with the multivariate structure constants

$$c_{\alpha\beta}^\gamma = \prod_{j=1}^{\infty} c_{\alpha_j \beta_j}^{\gamma_j}, \tag{B.11}$$

defined in terms of the univariate structure constants Eq. (B.9).

From this it is easy to see that

$$\mathbb{E}\left(H_\alpha H_\beta H_\gamma\right) = \mathbb{E}\left(H_\gamma \sum_\varepsilon c_{\alpha\beta}^\varepsilon H_\varepsilon\right) = c_{\alpha\beta}^\gamma \gamma!. \tag{B.12}$$

Products of more than two Hermite polynomials may be computed recursively, we here look at triple products as an example, using Eq. (B.10):

$$\begin{aligned} H_\alpha H_\beta H_\delta &= \left(\sum_\gamma c_{\alpha\beta}^\gamma H_\gamma\right) H_\delta \\ &= \sum_\varepsilon \left(\sum_\gamma c_{\gamma\delta}^\varepsilon c_{\alpha\beta}^\gamma\right) H_\varepsilon. \end{aligned} \tag{B.13}$$

B.4. The Hermite-Transform

A variant of the Hermite transform maps a random variable onto the set of expansion coefficients of the PCE [168]. Any random variable $r \in L_2(\Omega)$ which may be represented with a PCE

$$r(\omega) = \sum_{\alpha\in\mathcal{J}} r^\alpha H_\alpha(\theta(\omega)) \tag{B.14}$$

is mapped onto

$$\mathscr{H}(r) := (r^\alpha)_{\alpha\in\mathcal{J}} =: (r) \in \mathbb{R}^{\mathcal{J}}. \tag{B.15}$$

This way $\bar{r} := \mathbb{E}(r) = r^0$ and $\mathscr{H}(\bar{r}) = (r^0, 0, 0, ...)$, as well as $\tilde{r}(\omega) := r(\omega) - \bar{r}$ and $\mathscr{H}(\tilde{r}) = (0, (r^\alpha)_{\alpha\in\mathcal{J},\alpha>0})$.

These sequences may be seen also as the coefficients of power series in infinitely many complex variables $z \in \mathbb{C}^{\mathbb{N}}$, namely by

$$\sum_{\alpha\in\mathcal{J}} r^\alpha z^\alpha,$$

where $z^\alpha := \prod_j z_j^{\alpha_j}$. This is the original definition of the Hermite transform [168].

It can be used to easily compute the Hermite transform of the ordinary product like in Eq. (B.10), as

$$\mathscr{H}(H_\alpha H_\beta) = (c_{\alpha\beta}^\gamma)_{\gamma\in\mathcal{J}}. \tag{B.16}$$

With the structure constants Eq. (B.11) one defines the matrices $Q_2^\gamma := (c_{\alpha\beta}^\gamma)$ with indices α and β. With this notation the Hermite transform of the product of two random variables $r_1(\omega) = \sum_{\alpha\in\mathcal{J}} r_1^\alpha H_\alpha(\theta)$ and $r_2(\omega) = \sum_{\beta\in\mathcal{J}} r_2^\beta H_\beta(\theta)$ is

$$\mathscr{H}(r_1 r_2) = \left((r_1) Q_2^\gamma (r_2)^T)\right)_{\gamma\in\mathcal{J}}. \tag{B.17}$$

Each coefficient is a bilinear form in the coefficient sequences of the factors, and the collection of all those bilinear forms $\mathsf{Q}_2 = (Q_2^\gamma)_{\gamma\in\mathcal{J}}$ is a bilinear mapping that maps the coefficient sequences of r_1 and r_2 into the coefficient sequence of the product

$$\mathscr{H}(r_1 r_2) =: \mathsf{Q}_2((r_1),(r_2)) = \mathsf{Q}_2\left(\mathscr{H}(r_1), \mathscr{H}(r_2)\right). \tag{B.18}$$

Products of more than two random variables may now be defined recursively through the use of associativity. e.g. $r_1 r_2 r_3 r_4 = (((r_1 r_2) r_3) r_4)$:

$$\forall k > 2: \quad \mathscr{H}\left(\prod_{j=1}^k r_j\right) := \mathsf{Q}_k((r_1),(r_2),\ldots,(r_k)) := \\ \mathsf{Q}_{k-1}(\mathsf{Q}_2((r_1),(r_2)),(r_3)\ldots,(r_k)). \tag{B.19}$$

Each Q_k is again composed of a sequence of k-linear forms $\{Q_k^\gamma\}_{\gamma\in\mathcal{J}}$, which define each coefficient of the Hermite transform of the k-fold product.

B.5. Higher Order Moments of Hermite-Transforms

Consider RVs $\boldsymbol{r}_j(\omega) = \sum_{\alpha\in\mathcal{J}} \boldsymbol{r}_j^\alpha H_\alpha(\theta(\omega))$ with values in a vector space $\mathcal{V}$, then $\bar{\boldsymbol{r}}_j$, $\tilde{\boldsymbol{r}}_j(\omega)$, as well as $\boldsymbol{r}_j^\alpha$ are in $\mathcal{V}$. Any moment may be easily computed knowing the PCE. The k-th centred moment is defined as

$$\boldsymbol{M}^k_{\boldsymbol{r}_1\ldots\boldsymbol{r}_k} = \mathbb{E}\left(\otimes_{j=1}^k \tilde{\boldsymbol{r}}_j\right), \tag{B.20}$$

a tensor of order k. Thus it may be expressed via the PCE as

$$\boldsymbol{M}^k_{\boldsymbol{r}_1 \dots \boldsymbol{r}_k} = \sum_{\gamma^1, \dots, \gamma^k \neq 0} \mathbb{E}\left(\prod_{j=1}^{k} H_{\gamma^j}(\theta) \right) \otimes_{m=1}^{k} \boldsymbol{r}_m^{\gamma^m}, \tag{B.21}$$

and in particular:

$$\begin{aligned} \boldsymbol{C}_{\boldsymbol{r}_1 \boldsymbol{r}_2} &= \boldsymbol{M}^2_{\boldsymbol{r}_1 \boldsymbol{r}_2} = \mathbb{E}\left(\boldsymbol{r}_1 \otimes \boldsymbol{r}_2\right) \\ &= \sum_{\gamma, \beta > 0} \mathbb{E}\left(H_\gamma H_\beta\right) \boldsymbol{r}_1^\gamma \otimes \boldsymbol{r}_2^\beta \\ &= \sum_{\gamma > 0} \gamma! \, \boldsymbol{r}_1^\gamma \otimes \boldsymbol{r}_2^\gamma, \end{aligned} \tag{B.22}$$

as $\mathbb{E}\left(H_\gamma H_\beta\right) = \delta_{\gamma\beta}\gamma!$. The expected values of the products of Hermite polynomials in Eq. (B.21) may be computed analytically, by using the formulas from appendix B.3.

Appendix C.

Metrics for Comparison

In this appendix details are given on the comparison metrics used in the numerical examples.

C.1. Root Mean Square Error

The root mean square error of the estimator $\hat{x}(\omega)$ in relation to the the correct — or 'true' — value $\breve{x}$ is defined as

$$\mathrm{RMSE}(\hat{x}(\omega)) = \sqrt{\mathbb{E}\left((\hat{x} - \breve{x})^2\right)}. \tag{C.1}$$

If estimator and model are vectors of size n, the mean over the RMSE over all vector entries $\boldsymbol{x}(\omega) = (..., x_i, ...)^T$ is computed:

$$\mathrm{RMSE}(\hat{\boldsymbol{x}}(\omega)) = \frac{1}{n}\sum_{i=1}^{n}\sqrt{\mathbb{E}\left((\hat{x}_i - \breve{x}_i)^2\right)}. \tag{C.2}$$

C.2. Relative Error

The relative error of a functional $f(\cdot)$ of the estimator $\hat{x}(\omega)$ for the 'truth' $\breve{x}$ is defined for $\|\breve{x}\| \neq 0$ as

$$\mathrm{RE}(f(\hat{x})) = \frac{\|f(\hat{x}) - f(\breve{x})\|}{\|f(\breve{x})\|}, \tag{C.3}$$

where the norm $\|\cdot\|$ is chosen appropriately to the domain where $f(\cdot)$ and $\breve{x}$ reside. Commonly used functionals are the expectation $\mathbb{E}(\hat{x}) = \hat{\bar{x}}$, a median $p_{50}(\hat{x})$, or a mode of $p(\hat{x})$.

C.3. Sample Skewness and Kurtosis

The estimation of skewness from samples x_i drawn from the RV $x(\omega)$, with N being the sample size, can be performed according to

$$\operatorname{skew}(x) \approx \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^3}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2\right)^{3/2}} \tag{C.4}$$

with $\bar{x}$ being the sample mean.

Similarly, the kurtosis can be estimated according to

$$\operatorname{kurt}(x) \approx \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^4}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2\right)^{2}} - 3. \tag{C.5}$$

C.4. Recursive Computation of Mean and Variance Estimates from Monte Carlo Samples

During an estimation process, statistics of the estimator have to be computed for comparison purposes. As the estimator is regarded as an RV $r(\omega)$, the two most important statistical values are the mean

$$\bar{\boldsymbol{r}} := \mathbb{E}(\boldsymbol{r}) \tag{C.6}$$

and the (centralised) variance

$$\operatorname{Var}(\boldsymbol{r}) := \mathbb{E}\left((\boldsymbol{r} - \bar{\boldsymbol{r}})^2\right). \tag{C.7}$$

When the actual representation of the RV is a Monte Carlo sample — or there is a cheap method to obtain such a sample — it is convenient to estimate these values directly from the sample.

Given a set of samples of size N, $\{\boldsymbol{r}_1, ..., \boldsymbol{r}_N\}$, we can compute the sample mean as an estimation of the the true mean as

$$\bar{\boldsymbol{r}} \approx \bar{\boldsymbol{r}}_N := \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{r}_i \tag{C.8}$$

and the (centralised) sample variance as an unbiased estimation of the true variance as

$$\mathrm{Var}(\boldsymbol{r}) \approx \mathrm{Var}_N(\boldsymbol{r}) := \frac{1}{N-1} \sum_{i=1}^{N} (\boldsymbol{r}_i - \bar{\boldsymbol{r}})^{\otimes 2}. \tag{C.9}$$

However, it is often convenient to compute these estimators recursively or "on the fly" while new samples $\boldsymbol{r}_i$ of $\boldsymbol{r}(\omega)$, with $i \in \mathbb{N}$, arrive. This is especially true for the application of sequential methods as discussed in this work.

The estimators can be obtained recursively in the following way for the mean

$$\bar{\boldsymbol{r}}_1 = \boldsymbol{r}_1 \tag{C.10}$$

$$\bar{\boldsymbol{r}}_{i+1} = \bar{\boldsymbol{r}}_i + \frac{\boldsymbol{r}_{i+1} - \bar{\boldsymbol{r}}_i}{i+1} \tag{C.11}$$

and for the variance

$$\mathrm{Var}_1(\boldsymbol{r}) = 0 \tag{C.12}$$

$$\mathrm{Var}_{i+1}(\boldsymbol{r}) = \left(1 - \frac{1}{i}\right) \mathrm{Var}_i(\boldsymbol{r}) + (i+1)(\bar{\boldsymbol{r}}_{i+1} - \bar{\boldsymbol{r}}_i)^2. \tag{C.13}$$

See Weisstein [355] for a proof.

Appendix D.

Implementation of the Ensemble Kalman Filter

The EnKF implementation used for comparison in this work is described here for reference purposes.

There are mainly three important points to consider when implementing Eq. (2.49) using the stochastic sampling representation: for numerical stability one should not solve the set of normal equations $\boldsymbol{A} = (\boldsymbol{C_z} + \boldsymbol{C_\varepsilon})^{-1}(\boldsymbol{Y} - \boldsymbol{Z})$ straightforwardly. Due to possible ill conditioning, it is best solved by pseudo inversion, denoted by $(\cdot)^\dagger$, using a singular value decomposition (SVD) (e.g. Evensen [123]). For this, we replace $\boldsymbol{C_\varepsilon}$ with the ensemble estimate $\boldsymbol{C_y}$. This is an acceptable approximation (see Evensen [121], his section 3.4.3), which comes in handy now:

$$(\boldsymbol{C_z} + \boldsymbol{C_y})^\dagger = \left(\tilde{\boldsymbol{Z}}\,\tilde{\boldsymbol{Z}}^T + \tilde{\boldsymbol{Y}}\,\tilde{\boldsymbol{Y}}^T\right)^\dagger \tag{D.1}$$

which is, taking into account that $\tilde{\boldsymbol{Z}}\tilde{\boldsymbol{Y}}^T \equiv 0$ and $\tilde{\boldsymbol{Y}}\,\tilde{\boldsymbol{Z}}^T \equiv 0$,

$$= \left(\left(\tilde{\boldsymbol{Z}} + \tilde{\boldsymbol{Y}}\right)\left(\tilde{\boldsymbol{Z}} + \tilde{\boldsymbol{Y}}\right)^T\right)^\dagger, \tag{D.2}$$

with the SVD of one factor being

$$\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T = \tilde{\boldsymbol{Z}} + \tilde{\boldsymbol{Y}}. \tag{D.3}$$

Assume the singular values in the diagonal matrix $\boldsymbol{\Sigma}$ arranged descending

by size. Inserting Eq. (D.3) into Eq. (D.2) one obtains

$$\left(\left(\tilde{\boldsymbol{Z}}+\tilde{\boldsymbol{Y}}\right)\left(\tilde{\boldsymbol{Z}}+\tilde{\boldsymbol{Y}}\right)^{T}\right)^{\dagger}=\left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{T}\boldsymbol{U}^{T}\right)^{\dagger} \tag{D.4}$$

$$=\left(\boldsymbol{U}\boldsymbol{\Sigma}^{\dagger}\boldsymbol{\Sigma}^{\dagger^{T}}\boldsymbol{U}^{T}\right), \tag{D.5}$$

where $\boldsymbol{\Sigma}^{\dagger}$ is the pseudo-inverse of $\boldsymbol{\Sigma}$, a diagonal matrix with the inverse of the largest singular values from $\boldsymbol{\Sigma}$. Below a certain threshold in $\boldsymbol{\Sigma}$ the corresponding elements in $\boldsymbol{\Sigma}^{\dagger}$ are set to zero [142].

The second important point is that for the pseudo-inversion to work correctly, all components of the RVs $\boldsymbol{y}(\omega)$ and $\boldsymbol{z}(\omega)$ have to be on the same scale. Otherwise, in the pseudo-inversion, the singular vectors of small scale measurements are systematically associated with small singular values — which obviously are more easily truncated. Thus it could introduce a bias in the update towards large scale measurements. This can be easily fixed by scaling the RVs with the assumed measurement standard deviation, thus making them non-dimensional.

The third important point is that we use second order exact sampling in any case where a random number ensemble $\boldsymbol{\Xi}$ of size N is drawn from an n-dimensional standard normal distribution $\mathcal{N}(0,\boldsymbol{I}_n)$. This may be achieved by subtracting an eventual mean

$$\boldsymbol{\Xi} \leftarrow \boldsymbol{\Xi} - \bar{\boldsymbol{\xi}}\,\boldsymbol{1}_N^T$$

to ensure that each entry has exactly unit variance and no covariance between the entries exists:

$$\boldsymbol{\Xi} \leftarrow (\boldsymbol{C}_{\boldsymbol{\xi}}^{-1/2})\,\boldsymbol{\Xi}.$$

This may not be feasible for large n. There, an approximation to the above approach is to assume a strictly diagonal $\boldsymbol{C}_{\boldsymbol{\xi}}$ and correct only the standard deviation of every row $j = 1..n$

$$\forall j : (\boldsymbol{\Xi})_j \leftarrow (\boldsymbol{\Xi})_j \cdot \left(\sqrt{\mathrm{Var}((\boldsymbol{\Xi})_j)}\right)^{-1}$$

of the sample [122].

Related Publications

[1] O. Pajonk. Overview of system identification with focus on inverse modeling: Literature review. Informatikbericht 2009-03, Institut für Wissenschaftliches Rechnen, Technische Universität Braunschweig, June 2009. URL `http://www.digibib.tu-bs.de/?docid=00030245`.

[2] O. Pajonk, M. Krosche, R. Schulze-Riegert, R. Niekamp, and H. G. Matthies. Stochastic optimization using EA and EnKF: A comparison. In *Proceedings of the 11th European Conference on the Mathematics of Oil Recovery, September 8–11, Bergen, Norway*, 2008. URL `http://www.earthdoc.org/detail.php?pubid=23919`.

[3] O. Pajonk, R. Schulze-Riegert, and M. Krosche. Evolutionary optimization and stochastic filtering: A hybrid approach. In H. Rodrigues, C. Soares, J. Guedes, P. Fernandes, J. Folgado, C. Pina, and M. Neves, editors, *Proceedings of the 8th World Congress on Structural and Multidisciplinary Optimization, June 1–5, Lisboa, Portugal*. International Society for Structural and Multidisciplinary Optimization, 2009. ISBN: 987-989-20-1554-5.

[4] O. Pajonk, H. Mustafa, M. Krosche, and R. Schulze-Riegert. Ensemble based production optimization: A case example. In *DGMK-Tagungsbericht 2010-1*, 2010. ISBN 978-3-936418-05-0.

[5] O. Pajonk, R. Schulze-Riegert, M. Krosche, and H. G. Matthies. Multiscale wavelet analysis applied to localization of covariance estimates in EnKF. In *Proceedings of the 12th European Conference on the Mathematics of Oil Recovery, September 6–9, Oxford, United Kingdom*, 2010. URL `http://earthdoc.eage.org/detail.php?pubid=41337`.

[6] O. Pajonk, B. V. Rosić, A. Litvinenko, and H. G. Matthies. A

deterministic filter for non-Gaussian Bayesian estimation. Informatikbericht 2011-04, Institut für Wissenschaftliches Rechnen, Technische Universität Braunschweig, 2011. URL `http://www.digibib.tu-bs.de/?docid=00038994`.

[7] O. Pajonk, B. V. Rosić, A. Litvinenko, and H. G. Matthies. A deterministic filter for non-Gaussian state estimation. *Proceedings in Applied Mathematics and Mechanics*, 11(1):703–704, 2011. ISSN 1617-7061. doi:10.1002/pamm.201110341.

[8] O. Pajonk, B. V. Rosić, A. Litvinenko, and H. G. Matthies. A deterministic filter for non-Gaussian Bayesian estimation. *Physica D: Nonlinear Phenomena*, 241(7):775–788, April 2012. doi:10.1016/j.physd.2012.01.001.

[9] O. Pajonk, B. V. Rosić, and H. G. Matthies. Deterministic linear Bayesian updating of state and model parameters for a chaotic model. Informatikbericht 2012-01, Institut für Wissenschaftliches Rechnen, Technische Universität Braunschweig, 2012. URL `http://www.digibib.tu-bs.de/?docid=00042266`.

[10] B. V. Rosić, A. Litvinenko, O. Pajonk, and H. G. Matthies. Direct Bayesian update of polynomial chaos representations. Informatikbericht 2011-02, Institut für Wissenschaftliches Rechnen, Technische Universität Braunschweig, 2011. URL `http://www.digibib.tu-bs.de/?docid=00039000`.

[11] B. V. Rosić, H. G. Matthies, A. Litvinenko, O. Pajonk, A. Kucerova, and J. Sykora. Bayesian updating of uncertainties in the description of heat and moisture transport in heterogeneous materials. In *Proceedings of the V International Conference on Adaptive Modeling and Simulation (ADMOS), June 6–8, Paris, France*, 2011.

[12] R. Schulze-Riegert, M. Krosche, and O. Pajonk. Hybrid optimization coupling EnKF and evolutionary algorithms for history matching: A case example. In *Proceedings of the SPE EUROPEC/EAGE Annual Conference and Exhibition, June 8–11, Amsterdam, The Netherlands*, number SPE 121965, 2009. doi:10.2118/121965-MS.

[13] R. Schulze-Riegert, M. Krosche, and O. Pajonk. Data assimilation coupled to evolutionary algorithms: A case example in history match-

ing. In *Proceedings of the SPE/EAGE Reservoir Characterization and Simulation Conference, October 19–21, Abu Dhabi, UAE*, 2009. doi:10.2118/125512-MS.

Bibliography

[14] S. I. Aanonsen, G. Nævdal, D. S. Oliver, A. C. Reynolds, and B. Vallès. The ensemble Kalman filter in reservoir engineering - a review. *SPE Journal*, 14(3):393–412, September 2009. doi:10.2118/117274-PA.

[15] R. J. Adler and J. E. Taylor. *Random fields and geometry*. Springer monographs in mathematics. Springer-Verlag, Berlin, Germany, 2007. doi:10.1007/978-0-387-48116-6.

[16] A. Alcolea, J. Carrera, and A. Medina. Pilot points method incorporating prior information for solving the groundwater flow inverse problem. *Advances in Water Resources*, 29(11):1678–1689, November 2006. ISSN 0309-1708. doi:10.1016/j.advwatres.2005.12.009.

[17] D. Alspach and H. Sorenson. Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE Transactions on Automatic Control*, 17(4):439–448, Aug 1972. ISSN 0018-9286. URL `http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1100034`.

[18] B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Information and System Science Series. Prentice Hall, Englewood Cliffs, NJ, USA, 1979.

[19] J. L. Anderson. An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, 129:2884–2903, 2001. doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.

[20] J. L. Anderson. A local least squares framework for ensemble filtering. *Monthly Weather Review*, 131(4):634–642, April 2003. doi:10.1175/1520-0493(2003)131<0634:ALLSFF>2.0.CO;2.

[21] J. L. Anderson. An adaptive covariance inflation error correc-

tion algorithm for ensemble filters. *Tellus A*, 59(2):210–224, 2007. doi:10.1111/j.1600-0870.2006.00216.x.

[22] J. L. Anderson. Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus A*, 61(1):72–83, 2009. ISSN 1600-0870. doi:10.1111/j.1600-0870.2008.00361.x.

[23] J. L. Anderson. A non-Gaussian ensemble filter update for data assimilation. *Monthly Weather Review*, 138(11):4186–4198, 2010. doi:10.1175/2010MWR3253.1.

[24] J. L. Anderson and S. L. Anderson. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127(12):2741–2758, December 1999. doi:10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2.

[25] J. Anderson. Ensemble Kalman filters for large geophysical applications. *IEEE Control Systems Magazine*, 29(3):66–82, 2009. ISSN 0272-1708. doi:10.1109/MCS.2009.932222.

[26] A. Andrews. A square root formulation of the Kalman covariance equations. *AIAA Journal*, 6(6):1165–1166, June 1968. doi:10.2514/3.4696.

[27] D. Applebaum. *Probability and Information—An Integrated Approach*. Cambridge University Press, Cambridge, UK, second edition, 2008. URL `http://www.cambridge.org/9780521899048`.

[28] A. Apte, M. Hairer, A. Stuart, and J. Voss. Sampling the posterior: An approach to non-Gaussian data assimilation. *Physica D: Nonlinear Phenomena*, 230(1-2):50–64, June 2007. ISSN 0167-2789. doi:10.1016/j.physd.2006.06.009.

[29] I. Arasaratnam and S. Haykin. Cubature Kalman filters. *IEEE Transactions on Automatic Control*, 54(6):1254–1269, 2009. ISSN 0018-9286. doi:10.1109/TAC.2009.2019800.

[30] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Log-Euclidean metrics for fast and simple calculus on diffusion tensors.

Magn. Reson. Med., 56(2):411–421, 2006. ISSN 1522-2594. doi:10.1002/mrm.20965.

[31] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 29 (1):328–347, January 2007. doi:10.1137/050637996.

[32] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, Feb 2002. ISSN 1053-587X. doi:10.1109/78.978374.

[33] R. B. Ash. *Probability and Measure Theory*. Academic Press, New York, London, second edition, 2000.

[34] R. C. Aster, B. Borchers, and C. H. Thurber. *Parameter Estimation and Inverse Problems*, volume 90 of *International Geophysics Series*. Elsevier, Amsterdam, The Netherlands, 2005.

[35] A. Auger, S. Finck, N. Hansen, and R. Ros. BBOB 2010: Comparison tables of all algorithms on all noisy functions. Technical Report RT-389, INRIA, September 2010. URL `http://hal.inria.fr/inria-00516690/en/`.

[36] A. Bain and D. Crisan. *Fundamentals of Stochastic Filtering*, volume 60 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, Germany, 2009. doi:10.1007/978-0-387-76896-0.

[37] A. Banks, J. Vincent, and C. Anyakoha. A review of particle swarm optimization. part I: background and development. *Natural Computing*, 6(4):467–484, 2007. doi:10.1007/s11047-007-9049-5.

[38] A. Banks, J. Vincent, and C. Anyakoha. A review of particle swarm optimization. part II: hybridisation, combinatorial, multicriteria and constrained optimization, and indicative applications. *Natural Computing*, 7(1):109–124, March 2008. doi:10.1007/s11047-007-9050-z.

[39] R. Bannister. Localization in the EnKF - why does it increase the rank of p_f? Technical report, University of Reading, 2008.

[40] R. Baragona, F. Battaglia, and I. Poli. *Evolutionary Statistical Procedures: An Evolutionary Computation Approach to Statistical Procedures Designs and Applications*. Statistics and Computing. Springer-Verlag, Berlin, Germany, 2011. doi:10.1007/978-3-642-16218-3.

[41] A. Bashi, V. Jilkov, X. Li, and H. Chen. Distributed implementations of particle filters. *Proceedings of the Sixth International Conference of Information Fusion 2003*, 2:1164–1171, 2003.

[42] M. Bayes and M. Price. An essay towards solving a problem in the doctrine of chances. by the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philosophical Transactions*, 53:370–418, 1763. doi:10.1098/rstl.1763.0053.

[43] J. L. Beck and K. M. Zuev. Asymptotically independent Markov sampling: a new MCMC scheme for Bayesian inference. 2011, `arXiv:1110.1880`.

[44] J. Behrens and R. LeVeque. Modeling and simulating tsunamis with an eye to hazard mitigation. *SIAM News*, 44(4):1–8, May 2011.

[45] R. E. Bellman. *Dynamic Programming*. Princeton University Press, Princeton, NJ, 1957.

[46] A. Ben-Naim. *Statistical Thermodynamics Based on Information—A Farewell to Entropy*. World Scientific Publishing Co. Pte. Ltd., Singapore, 2008.

[47] T. Bengtsson, C. Snyder, and D. Nychka. Toward a nonlinear ensemble filter for high-dimensional systems. *Journal of Geophysical Research*, 108:–, December 2003. doi:10.1029/2002JD002900.

[48] A. F. Bennett. *Inverse Modeling of the Ocean and Atmosphere*. Cambridge University Press, Cambridge, UK, 2002.

[49] P. R. Benson. ISO 8000 data quality – the fundamentals part 1. *Real-World Decision Support Journal*, 3(4):1–4, 2009.

[50] K. Bergemann and S. Reich. A localization technique for ensemble

Kalman filters. *Quarterly Journal of the Royal Meteorological Society*, 136(648):701–707, 2010. ISSN 1477-870X. doi:10.1002/qj.591.

[51] K. Bergemann, G. Gottwald, and S. Reich. Ensemble propagation and continuous matrix factorization algorithms. *Quarterly Journal of the Royal Meteorological Society*, 135(643):1560–1572, 2009. ISSN 1477-870X. doi:10.1002/qj.457.

[52] K. Bergemann and S. Reich. A mollified ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 136(651): 1636–1643, 2010. ISSN 1477-870X. doi:10.1002/qj.672.

[53] N. Bergman, L. Ljung, and F. Gustafsson. Terrain navigation using Bayesian statistics. *IEEE Control Systems Magazine*, 19(3):33–40, 1999. ISSN 1066-033X. doi:10.1109/37.768538.

[54] L. Berre and G. Desroziers. Filtering of background error variances and correlations by local spatial averaging: A review. *Monthly Weather Review*, 138(10):3693–3720, October 2010. doi:10.1175/2010MWR3111.1.

[55] D. P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, Nashua, NH, USA, third edition, 2005.

[56] D. P. Bertsekas and S. E. Shreve. *Stochastic Optimal Control: The Discrete-Time Case*, volume 5 of *Optimization and Neural Computation Series*. Athena Scientific, Nashua, NH, USA, 1996.

[57] H.-G. Beyer. *The Theory of Evolution Strategies*. Natural Computing. Springer-Verlag, Berlin, Germany, 2001.

[58] H.-G. Beyer and H.-P. Schwefel. Evolution strategies – a comprehensive introduction. *Natural Computing*, 1(1):3–52, March 2002. doi:10.1023/A:1015059928466.

[59] A. Bianco, A. Cominelli, L. Dovera, G. Nævdal, and B. Valles. History matching and production forecast uncertainty by means of the ensemble Kalman filter – a real field application. In *Proceedings of the SPE Europec/EAGE Annual Conference and Exhibition, London, United Kingdom, 11–14 June*, 2007. URL `http://www.earthdoc.org/detail.php?pubid=6771`.

[60] L. Biegler, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B. Mallick, L. Tenorio, B. van Bloemen Waanders, K. Willcox, and Y. Marzouk, editors. *Large-Scale Inverse Problems and Quantification of Uncertainty*. Wiley Series in Computational Statistics. John Wiley & Sons, Chichester, 2011. URL `http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470697431.html`.

[61] G. J. Bierman, editor. *Factorization Methods for Discrete Sequential Estimation*, volume 128 of *Mathematics in Science and Engineering*. Academic Press, Inc., New York, 1977.

[62] P. Billingsley. *Probability and Measure*. John Wiley & Sons, Chichester, third edition, 1995.

[63] C. H. Bishop and D. Hodyss. Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 133 (629):2029–2044, 2007. doi:10.1002/qj.169.

[64] C. H. Bishop and D. Hodyss. Ensemble covariances adaptively localized with ECO-RAP. part 1: tests on simple error models. *Tellus A*, 61(1):84–96, 2009. ISSN 1600-0870. doi:10.1111/j.1600-0870.2008.00371.x.

[65] C. H. Bishop, B. J. Etherton, and S. J. Majumdar. Adaptive sampling with the ensemble transform Kalman filter. part I: Theoretical aspects. *Monthly Weather Review*, 129(3):420–436, March 2001. doi:10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.

[66] E. D. Blanchard, A. Sandu, and C. Sandu. A polynomial chaos-based Kalman filter approach for parameter estimation of mechanical systems. *Journal of Dynamic Systems, Measurement, and Control*, 132(6):061404, 2010. doi:10.1115/1.4002481.

[67] E. D. Blanchard. *Polynomial Chaos Approaches to Parameter Estimation and Control Design for Mechanical Systems with Uncertain Parameters*. PhD thesis, Department of Mechanical Engineering, VirginiaTech University, 2010. URL `http://scholar.lib.vt.edu/theses/available/etd-04092010-000752/`.

[68] A. Bobrowski. *Functional Analysis for Probability and Stochastic*

Processes. Cambridge University Press, Cambridge, UK, 2005. URL http://www.cambridge.org/9780521831666.

[69] V. I. Bogachev. *Gaussian Measures*. Number 62 in Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, USA, 1998.

[70] V. I. Bogachev. *Measure Theory: Volume II*. Springer-Verlag, Berlin, Germany, 2007.

[71] D. Bosq and D. Blanke. *Inference and Prediction in Large Dimensions*. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester, 2007.

[72] Z. I. Botev, J. F. Grotowski, and D. P. Kroese. Kernel density estimation via diffusion. *Annals of Statistics*, 38(5):2916–2957, 2010. doi:10.1214/10-AOS799.

[73] N. Bourbaki. *Topological vector spaces*, volume 1. Springer-Verlag, Berlin, Germany, 1987.

[74] J. J. Buckley. *Fuzzy Probabilities—New Approach and Applications*. Springer-Verlag, Berlin, Germany, 2005.

[75] J. J. Buckley. *Fuzzy Probability and Statistics*, volume 196 of *Studies in Fuzziness and Soft Computing*. Springer-Verlag, Berlin, Germany, 2006.

[76] R. Bucy. Linear and nonlinear filtering. *Proceedings of the IEEE*, 58 (6):854–864, 1970. ISSN 0018-9219. doi:10.1109/PROC.1970.7792.

[77] A. Budhiraja, L. Chen, and C. Lee. A survey of numerical methods for nonlinear filtering problems. *Physica D: Nonlinear Phenomena*, 230(1-2):27–36, June 2007. ISSN 0167-2789. doi:10.1016/j.physd.2006.08.015.

[78] G. Burgers, P. J. van Leeuwen, and G. Evensen. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, 126:1719–1724, 1998. doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.

[79] J. C. Butcher. *Numerical Methods for Ordinary Differential Equations*. John Wiley & Sons, Chichester, second edition, 2008.

[80] T. Bäck. *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*. Oxford University Press, Oxford, UK, 1996.

[81] J. Caers and T. Zhang. Multiple-point geostatistics: a quantitative vehicle for integrating geologic analogs into multiple reservoir models. In G. M. Grammer, P. M. Harris, and G. P. Eberli, editors, *Integration of outcrop and modern analog data in reservoir models: AAPG Memoir*. American Association of Petroleum Geologists, Alexandria, VA, 2004.

[82] R. H. Cameron and W. T. Martin. Transformations of Wiener integrals under a general class of linear transformations. *Transactions of the American Mathematical Society*, 58(2):184–219, September 1945. URL `http://www.jstor.org/stable/1990282`.

[83] R. H. Cameron and W. T. Martin. The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals. *The Annals of Mathematics*, 48(2):385–392, 1947. URL `http://www.jstor.org/stable/1969178`.

[84] L. Chambers, editor. *Practical Handbook of Genetic Algorithms: Applications*, volume 1. CRC Press, Boca Raton, FL, 1999.

[85] L. Chambers, editor. *Practical Handbook of Genetic Algorithms: New Frontiers*, volume 2. CRC Press, Boca Raton, FL, 1999.

[86] Y. Chen. *Ensemble-based Closed Loop Production Optimization*. PhD thesis, University of Oklahoma, 2008.

[87] Y. Chen and D. S. Oliver. Ensemble-based closed-loop optimization on Brugge field. In *SPE Reservoir Simulation Symposium, The Woodlands, Texas, USA*, 2009. doi:10.2118/118926-MS. SPE118926.

[88] Y. Chen, D. S. Oliver, and D. Zhang. Data assimilation for nonlinear problems by ensemble Kalman filter with reparameteriza-

tion. *Journal of Petroleum Science and Engineering*, 66:1–14, 2009. doi:10.1016/j.petrol.2008.12.002.

[89] Y. Chen, S. Nourazari, and D. S. Oliver. Multiscale parameterization with adaptive localization for improved assimilation of time-lapse data. In *Proceedings of the SPE Reservoir Simulation Symposium, The Woodlands, Texas, U.S.A., 21–23 February*, 2011.

[90] Z. Chen. Bayesian filtering: From Kalman filters to particle filters, and beyond. Technical report, Adaptive Systems Lab., McMaster University, Hamilton, ON, Canada, 2003. URL `http://soma.crl.mcmaster.ca/~zhechen/download/ieee_bayesian.ps.gz`. Communications Research Laboratory, McMaster University.

[91] S. Chib and E. Greenberg. Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4):327–335, 1995. doi:10.2307/2684568.

[92] S. E. Cohn, A. da Silva, J. Guo, M. Sienkiewicz, and D. Lamich. Assessing the effects of data selection with the DAO physical-space statistical analysis system. *Monthly Weather Review*, 126(11):2913–2926, 1998. doi:10.1175/1520-0493(1998)126<2913:ATEODS>2.0.CO;2.

[93] S. L. Cotter, M. Dashti, and A. M. Stuart. Approximation of Bayesian inverse problems for PDEs. *SIAM Journal on Numerical Analysis*, 48(1):322–345, 2010. doi:10.1137/090770734.

[94] G. da Prato. *An Introduction to Infinite-Dimensional Analysis*. Springer-Verlag, Berlin, Germany, 2006.

[95] S. Da Veiga and V. Gervais. Local adaptive parameterization for the history matching of 3D seismic data. *Computational Geosciences*, Online First:1–16, 2011. ISSN 1420-0597. doi:10.1007/s10596-011-9241-7.

[96] F. Daum. Nonlinear filters: beyond the Kalman filter. *IEEE Aerospace and Electronic Systems Magazine*, 20(8):57–69, 2005. ISSN 0885-8985. doi:10.1109/MAES.2005.1499276.

[97] F. Daum and J. Huang. Particle flow for nonlinear filters. In

Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 5920–5923, 2011. doi:10.1109/ICASSP.2011.5947709.

[98] B. J. Debusschere, H. N. Najm, P. P. Pebay, O. M. Knio, R. G. Ghanem, and O. P. Le Maitre. Numerical challenges in the use of polynomial chaos representations for stochastic processes. *SIAM Journal on Scientific Computing*, 26(2):698–719, January 2004. doi:10.1137/S1064827503427741.

[99] A. Deckmyn and L. Berre. A wavelet approach to representing background error covariances in a limited-area model. *Monthly Weather Review*, 133(5):1279–1294, May 2005. doi:10.1175/MWR2929.1.

[100] J. E. Dennis and J. J. Moré. Quasi-Newton methods, motivation and theory. *SIAM Review*, 19(1):46–89, 1977. URL `http://www.jstor.org/stable/2029325`.

[101] G. Desroziers, L. Berre, B. Chapnik, and P. Poli. Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, 131 (613):3385–3396, 2005. ISSN 1477-870X. doi:10.1256/qj.05.108.

[102] C. V. Deutsch and A. G. Journel. *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press, Oxford, UK, second edition, 1997.

[103] P. J. Diggle and P. J. Ribeiro. *Model-based Geostatistics*. Springer Series in Statistics. Springer-Verlag, Berlin, Germany, 2007.

[104] A. Doucet and V. Tadic. Parameter estimation in general state-space models using particle methods. *Annals of the Institute of Statistical Mathematics*, 55(2):409–422, June 2003. doi:10.1007/BF02530508.

[105] A. Doucet, N. D. Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science. Springer-Verlag, Berlin, Germany, 2001.

[106] L. Dovera and E. Della Rossa. Multimodal ensemble Kalman filter-

ing using Gaussian mixture models. *Computational Geosciences*, 15 (2):307–323, 2011. ISSN 1420-0597. doi:10.1007/s10596-010-9205-3.

[107] L. Dovera and E. Della Rossa. Improved initial ensemble generation coupled with ensemble square root filters and inflation to estimate uncertainty. *Computational Geosciences*, Online First:1–17, 2011. ISSN 1420-0597. doi:10.1007/s10596-011-9243-5.

[108] K. Doya, S. Ishii, A. Pouget, and R. P. N. Rao, editors. *Bayesian Brain: Probabilistic Approaches to Neural Coding*. Computational Neuroscience. The MIT Press, Cambridge, MA, USA, 2007.

[109] M. Drugan and D. Thierens. Evolutionary Markov chain Monte Carlo. In P. Liardet, P. Collet, C. Fonlupt, E. Lutton, and M. Schoenauer, editors, *Artificial Evolution, 6th International Conference, Evolution Artificielle, EA 2003, Marseilles, France, October 27-30, 2003, Revised Selected Papers*, volume 2936 of *Lecture Notes in Computer Science*, pages 63–76. Springer Berlin / Heidelberg, 2004. doi:10.1007/978-3-540-24621-3_6.

[110] T. A. El Moselhy and Y. M. Marzouk. Bayesian inference with optimal maps. *Journal of Computational Physics*, Accepted for publication, September 2011, `arXiv:1109.1516`. Submitted for Publication.

[111] T. A. El Moselhy. *Field Solver Technologies for Variation-Aware Interconnect Parasitic Extraction*. PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, February 2010.

[112] J. Elsrodt. *Maß- und Integrationstheorie*. Springer-Lehrbuch. Springer-Verlag, Berlin, Germany, fourth edition, 2005. doi:10.1007/b137969.

[113] A. Emerick and A. Reynolds. Combining sensitivities and prior information for covariance localization in the ensemble Kalman filter for petroleum reservoir applications. *Computational Geosciences*, 15 (2):251–269, 2011. ISSN 1420-0597. doi:10.1007/s10596-010-9198-y.

[114] X. Emery and M. Peláez. Assessing the accuracy of sequential Gaus-

sian simulation and cosimulation. *Computational Geosciences*, 15 (4):673–689, 2011. ISSN 1420-0597. doi:10.1007/s10596-011-9235-5.

[115] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[116] Y. Ephraim and N. Merhav. Hidden Markov processes. *IEEE Transactions on Information Theory*, 48(6):1518 –1569, June 2002. ISSN 0018-9448. doi:10.1109/TIT.2002.1003838.

[117] O. G. Ernst, A. Mugler, H.-J. Starkloff, and E. Ullmann. On the convergence of generalized polynomial chaos expansions. Technical Report 60, DFG-SPP 1324, August 2010.

[118] G. Evensen, J. Hove, H. C. Meisingset, E. Reiso, K. S. Seim, and O. Espelid. Using the EnKF for assisted history matching of a North Sea reservoir model. In *Proceedings of the SPE Reservoir Simulation Symposium, Woodlands, Texas, 26.–28. February*, number SPE-106184, 2007. doi:10.2118/106184-MS.

[119] G. Evensen. Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *Journal of Geophysical Research*, 97(C11):17,905–17,924, 1992. doi:10.1029/92JC01972.

[120] G. Evensen. Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5):10,143–10,162, 1994. doi:10.1029/94JC00572.

[121] G. Evensen. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367, November 2003. doi:10.1007/s10236-003-0036-9.

[122] G. Evensen. Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics*, 54(6):539–560, December 2004. doi:10.1007/s10236-004-0099-2.

[123] G. Evensen. *Data Assimilation — The Ensemble Kalman Filter*. Springer-Verlag, Berlin, Germany, second edition, 2009.

[124] G. Evensen. The ensemble Kalman filter for combined state and parameter estimation. *IEEE Control Systems Magazine*, 29:82–104, 2009. doi:10.1109/MCS.2009.932223.

[125] G. Evensen and P. J. van Leeuwen. An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Review*, 128(6):1852–1867, 2000. doi:10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2.

[126] J. Field, Richard V. and M. D. Grigoriu. Convergence properties of polynomial chaos approximations for L_2 random variables. Technical Report SAND2007-1262, Sandia National Laboratories, 2007.

[127] R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222(594-604):309–368, 1922. doi:10.1098/rsta.1922.0009.

[128] A. M. Fraser. *Hidden Markov Models and Dynamical Systems*. Society for Industrial and Applied Mathematics, Philadelphia, USA, 2008.

[129] R. Frič and M. Papčo. A categorical approach to probability theory. *Studia Logica*, 94(2):215–230, 2010. ISSN 0039-3215. doi:10.1007/s11225-010-9232-z.

[130] M. Fujisaki, G. Kallianpur, and H. Kunita. Stochastic differential equations for the non linear filtering problem. *Osaka Journal of Mathematics*, 9(1):19–40, 1972. URL `http://projecteuclid.org/euclid.ojm/1200693535`.

[131] I. Fukumori. A partitioned Kalman filter and smoother. *Monthly Weather Review*, 130(5):1370–1383, 2002. doi:10.1175/1520-0493(2002)130<1370:APKFAS>2.0.CO;2.

[132] R. Furrer, M. G. Genton, and D. Nychka. Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15(3):502–523, September 2006. ISSN 1061-8600. doi:10.1198/106186006X132178.

[133] C. Gaetan and X. Guyon. *Spatial Statistics and Modeling*.

Springer Series in Statistics. Springer-Verlag, Berlin, Germany, 2010. doi:10.1007/978-0-387-92257-7.

[134] G. Gaspari and S. E. Cohn. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554):723–757, 1999. doi:10.1002/qj.49712555417.

[135] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton, FL, USA, second edition, 2004.

[136] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–742, 1984. doi:10.1109/TPAMI.1984.4767596.

[137] Z. Ghahramani. An introduction to hidden Markov models and Bayesian networks. *Journal of Pattern Recognition and Artificial Intelligence*, 15(1):9–42, 2001.

[138] R. G. Ghanem and P. D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Dover Publications, Inc., Mineola, NY, 2003. Revised edition of the 1991 book of Springer Verlag, New York.

[139] V. Ginting, F. Pereira, M. Presho, and S. Wo. Application of the two-stage Markov chain Monte Carlo method for characterization of fractured reservoirs using a surrogate flow model. *Computational Geosciences*, 17:1–17, 2011. ISSN 1420-0597. doi:10.1007/s10596-011-9236-4.

[140] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Boston, MA, USA, 1989. ISBN 0201157675. URL `http://portal.acm.org/citation.cfm?id=534133`.

[141] M. Goldstein and D. Wooff. *Bayesian Linear Statistics—Theory and Methods*. John Wiley & Sons, Chichester, 2007.

[142] G. H. Golub and C. F. van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, third edition, 1996.

[143] N. Gordon, D. Salmond, and A. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings F: Radar and Signal Processing*, volume 140, pages 107–113, April 1993. URL `http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=210672`.

[144] C. Gourieroux and A. Monfort. *Statistics and Econometric Models: General Concepts, Estimation, Prediction and Algorithms*, volume 1. Cambridge University Press, Cambridge, UK, 1995.

[145] M. Grabe. *Measurement Uncertainties in Science and Technology*. Springer-Verlag, Berlin, Germany, 2005.

[146] R. M. Gray. *Entropy and Information Theory*. Springer-Verlag, Berlin, Germany, second edition, 2011. doi:10.1007/978-1-4419-7970-4.

[147] Y. Gu and D. S. Oliver. An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. *SPE Journal*, 12(4):438–446, November 2007. doi:10.2118/108438-PA.

[148] W. W. Hager. Updating the inverse of a matrix. *SIAM Review*, 31 (2):221–239, 1989. doi:10.1137/1031049.

[149] J. Y. Halpern. *Reasoning about Uncertainty*. The MIT Press, Cambridge, MA, USA, 2003.

[150] T. M. Hamill and C. Snyder. A hybrid ensemble Kalman filter-3D variational analysis scheme. *Monthly Weather Review*, 128(8):2905–2919, 2000. doi:10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2.

[151] T. M. Hamill, J. S. Whitaker, and C. Snyder. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, 129(11):2776–2790, November 2001. doi:10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2.

[152] L. P. Hansen and T. J. Sargent. *Rational Expectations Econometrics*. Westview Press, Boulder, San Francisco, Oxford, 1991.

[153] N. Hansen. Updated list of CMA-ES applications. On-

line. URL http://www.bionik.tu-berlin.de/user/niko/cmaapplications.pdf.

[154] N. Hansen. *Towards a New Evolutionary Computation*, volume 192 of *Studies in Fuzziness and Soft Computing*, chapter The CMA Evolution Strategy: A Comparing Review, pages 75–102. Springer, 2006. doi:10.1007/3-540-32494-1_4.

[155] N. Hansen. The CMA evolution strategy: A tutorial. Technical report, Centre de recherche INRIA Saclay – Île-de-France, 2009.

[156] N. Hansen and S. Kern. Evaluating the CMA evolution strategy on multimodal test functions. In X. Yao, E. Burke, J. Lozano, J. Smith, J. Merelo-Guervós, J. Bullinaria, J. Rowe, P. Tino, A. Kabán, and H.-P. Schwefel, editors, *Lecture Notes in Computer Science*, volume 3242, pages 282–291. Springer, Berlin, Heidelberg, 2004. doi:10.1007/978-3-540-30217-9_29.

[157] J. A. Hartigan. Linear bayesian methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, 31(3):446–454, 1969. URL http://www.jstor.org/stable/2984348.

[158] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, April 1970. doi:10.1093/biomet/57.1.97.

[159] V. Haugen, L.-J. Natvik, G. Evensen, A. Berg, K. Flornes, and G. Nævdal. History matching using the ensemble Kalman filter on a North Sea field case. In *Proceedings of the SPE Annual Technical Conference and Exhibition, September 24–27, San Antonio, Texas, U.S.A*, 2006. SPE-102430.

[160] A. W. Heemink, M. Verlaan, and A. J. Segers. Variance reduced ensemble Kalman filtering. *Monthly Weather Review*, 129(7):1718–1728, July 2001. doi:10.1175/1520-0493(2001)129<1718:VREKF>2.0.CO;2.

[161] V. Heidrich-Meisner and C. Igel. Evolution strategies for direct policy search. In G. Rudolph, T. Jansen, and S. Lucas, editors, *Parallel Problem Solving from Nature (PPSN X)*, volume 5199 of

Lecture Notes in Computer Science, pages 428–437. Springer-Verlag, Berlin, Germany, 2008.

[162] V. Heidrich-Meisner and C. Igel. Uncertainty handling CMA-ES for reinforcement learning. In *Proceedings of the 11th Annual conference on Genetic and evolutionary computation*, 2009. doi:10.1145/1569901.1570064.

[163] T. Hida and S. Si. *Lectures on white noise functionals*. World Scientific Publishing Co. Pte. Ltd., Singapore, 2008.

[164] T. Hida, H.-H. Kuo, J. Potthoff, and L. Streit. *White Noise—An Infinite Dimensional Calculus*. Kluwer, Dordrecht, 1999.

[165] T. Higuchi. Genetic algorithm and Monte Carlo filter. In *Proceedings of the Institute of Statistical Mathematics*, volume 44, pages 19–30, 1996. URL `http://www.ism.ac.jp/editsec/toukei/tokeisuri-44-1e.html`. Japanese language.

[166] T. Higuchi. Monte Carlo filter using the genetic algorithm operators. *Journal of Statistical Computation and Simulation*, 59(1):1–23, 1997. ISSN 0094-9655. doi:10.1080/00949659708811843.

[167] T. Higuchi. Self-organizing time series model. In A. Doucet, J. de Freitas, and N. Gordon, editors, *Sequential Monte Carlo Methods in Practice*, pages 429–444. Springer-Verlag, Berlin, Germany, 2001.

[168] H. Holden, B. Øksendal, J. Ubøe, and T. Zhang. *Stochastic Partial Differential Equations: A Modeling, White Noise Functional Approach*. Universitext. Springer-Verlag, Berlin, Germany, second edition, 2010. doi:10.1007/978-0-387-89488-1.

[169] I. Hoteit, D.-T. Pham, G. Triantafyllou, and G. Korres. A new approximate solution of the optimal nonlinear filter for data assimilation in meteorology and oceanography. *Monthly Weather Review*, 136(1):317–334, January 2008. ISSN 0027-0644. doi:10.1175/2007MWR1927.1.

[170] P. L. Houtekamer and H. L. Mitchell. Ensemble Kalman filter-

ing. *Quarterly Journal of the Royal Meteorological Society*, 131(613): 3269–3289, 2005. ISSN 1477-870X. doi:10.1256/qj.05.135.

[171] P. L. Houtekamer and H. L. Mitchell. Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126:796–811, March 1998. doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

[172] P. L. Houtekamer and H. L. Mitchell. A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129(1):123–137, January 2001. doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

[173] P. L. Houtekamer, H. L. Mitchell, and X. Deng. Model error representation in an operational ensemble Kalman filter. *Monthly Weather Review*, 137(7):2126–2143, July 2009. doi:10.1175/2008MWR2737.1.

[174] L. Hu, G. Blanc, and B. Noetinger. Gradual deformation and iterative calibration of sequential stochastic simulations. *Mathematical Geology*, 33(4):475–489, 2001. ISSN 0882-8121. doi:10.1023/A:1011088913233.

[175] B. R. Hunt, E. Kalnay, E. J. Kostelich, E. Ott, D. J. Patil, T. Sauer, I. Szunyogh, J. A. Yorke, and A. V. Zimin. Four-dimensional ensemble Kalman filtering. *Tellus A*, 56(4):273–277, 2004. ISSN 1600-0870. doi:10.1111/j.1600-0870.2004.00066.x.

[176] B. R. Hunt, E. J. Kostelich, and I. Szunyogh. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, 230(1-2):112–126, June 2007. ISSN 0167-2789. doi:10.1016/j.physd.2006.11.008.

[177] O. C. Ibe. *Markov Processes for Stochastic Modeling*. Academic Press, New York, London, 2009.

[178] International Standardisation Organisation. ISO 80000-2: Mathematical signs and symbols to be used in the natural sciences and technology. Electronically, 2009. URL `http://www.iso.org/iso/catalogue_detail?csnumber=31887`.

[179] K. Ito and K. Xiong. Gaussian filters for nonlinear filtering problems. *IEEE Transactions on Automatic Control*, 45(5):910–927, May 2000.

[180] A. J. Izenman. *Modern Multivariate Statistical Techniques—Regression, Classification, and Manifold Learning*. Springer Texts in Statistics. Springer-Verlag, Berlin, Germany, 2008. doi:10.1007/978-0-387-78189-1.

[181] B. Jafarpour and M. Khodabakhshi. A probability conditioning method (PCM) for integration of production data into pattern-based facies simulation. In *Proceedings of the 12th European Conference on the Mathematics of Oil Recovery, Oxford, United Kingdom*, 2010.

[182] J. D. Jakeman and S. G. Roberts. Stochastic Galerkin and collocation methods for quantifying uncertainty in differential equations: a review. In G. N. Mercer and A. J. Roberts, editors, *Proceedings of the 14th Biennial Computational Techniques and Applications Conference, CTAC-2008*, volume 50 of *ANZIAM J.*, pages C815–C830, February 2009. URL `http://anziamj.austms.org.au/ojs/index.php/ANZIAMJ/article/view/1410`.

[183] T. Janjić, L. Nerger, A. Albertella, J. Schröter, and S. Skachko. On domain localization in ensemble-based Kalman filter algorithms. *Monthly Weather Review*, 139(7):2046–2060, March 2011. ISSN 0027-0644. doi:10.1175/2011MWR3552.1.

[184] S. Janson. *Gaussian Hilbert spaces*, volume 129 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, UK, 1997.

[185] E. T. Jaynes. *Maximum-Entropy and Bayesian Methods in Science and Engineering*, chapter How Does the Brain do Plausible Reasoning?, pages 1–24. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1988.

[186] E. T. Jaynes. *Probability Theory — The Logic of Science*. Cambridge University Press, Cambridge, UK, 2003.

[187] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, Inc., New York, 1970.

[188] F. V. Jensen and T. D. Nielsen. *Bayesian Networks and Decision Graphs*. Information Science and Statistics. Springer-Verlag, Berlin, Germany, 2007.

[189] C. Johns and J. Mandel. A two-stage ensemble Kalman filter for smooth data assimilation. *Environmental and Ecological Statistics*, 15(1):101–110, March 2008. doi:10.1007/s10651-007-0033-0.

[190] S. Julier and J. Uhlmann. A new extension of the Kalman filter to nonlinear systems. In *SPIE AeroSense Symposium, April 21-24*, 1997.

[191] S. Julier and J. K. Uhlmann. A general method for approximating nonlinear transformations of probability distributions. Technical report, University of Oxford, 1996.

[192] S. Julier and J. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, Mar 2004. ISSN 0018-9219. doi:10.1109/JPROC.2003.823141.

[193] S. Julier and J. Uhlmann. Correction to "unscented filtering and nonlinear estimation". *Proceedings of the IEEE*, 92(3):401–422, Mar 2004. ISSN 0018-9219. doi:10.1109/JPROC.2003.823141.

[194] S. Julier, J. Uhlmann, and H. Durrant-Whyte. A new approach for filtering nonlinear systems. In *Proceedings of the American Control Conference, 1995*, volume 3, pages 1628–1632, 1995. doi:10.1109/ACC.1995.529783.

[195] K. Jänich. *Topologie*. Springer-Lehrbuch. Springer-Verlag, Berlin, Germany, eigth edition, 2005. doi:10.1007/b138142.

[196] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1-2):99–134, May 1998. ISSN 0004-3702. doi:10.1016/S0004-3702(98)00023-X.

[197] C. Kalender and A. Schöttl. Nonlinear filtering using sparse grids. In F. Holzapfel and S. Theil, editors, *Advances in Aerospace Guidance, Navigation and Control*, pages 359–371. Springer Berlin Heidelberg, 2011. doi:10.1007/978-3-642-19817-5_28.

[198] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transaction of the ASME - Journal of Basic Engineering*, 82(Series D):35–45, March 1960. URL `http://www.cs.unc.edu/~welch/kalman/media/pdf/Kalman1960.pdf`.

[199] E. Kalnay, H. Li, T. Miyoshi, S.-C. Yang, and J. Ballabrera-Poy. Response to the discussion on "4-D-Var or EnKF?" by Nils Gustafsson. *Tellus A*, 59(5):778–780, 2007. ISSN 1600-0870. doi:10.1111/j.1600-0870.2007.00263.x.

[200] K. Kastella and C. Kreucher. Multiple model nonlinear filtering for low signal ground target applications. *IEEE Transactions on Aerospace and Electronic Systems*, 41(2):549–564, 2005. ISSN 0018-9251. doi:10.1109/TAES.2005.1468747.

[201] A. Keese. *Numerical Solution of Systems with Stochastic Uncertainties—A General Purpose Framework for Stochastic Finite Elements*. PhD thesis, Fachbereich für Mathematik und Informatik, Technische Universität Braunschweig, 2004. URL `http://www.digibib.tu-bs.de/?docid=00001595`.

[202] A. Keese and H. G. Matthies. Sparse quadrature as an alternative to Monte Carlo for stochastic finite element techniques. *PAMM*, 3 (1):493–494, 2003. doi:10.1002/pamm.200310516.

[203] A. Keese and H. G. Matthies. Numerical methods and Smolyak quadrature for nonlinear stochastic partial differential equations. Informatikbericht 2003-5, Institut für Wissenschaftliches Rechnen, Technische Universität Braunschweig, 2003. URL `http://www.digibib.tu-bs.de/?docid=00001471`.

[204] J. Kennedy and R. Eberhart. Particle swarm optimization. *Neural Networks, 1995. Proceedings., IEEE International Conference on*, 4:1942–1948 vol.4, Nov/Dec 1995. doi:10.1109/ICNN.1995.488968.

[205] J. D. Kepert. On ensemble representation of the observation-error covariance in the ensemble Kalman filter. *Ocean Dynamics*, 54(6): 561–569, December 2004. doi:10.1007/s10236-004-0104-9.

[206] J. D. Kepert. Covariance localisation and balance in an ensemble

kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 135(642):1157–1176, 2009. ISSN 1477-870X. doi:10.1002/qj.443.

[207] J. Kepert. Balance-aware covariance localisation for atmospheric and oceanic ensemble Kalman filters. *Computational Geosciences*, 15(2):239–250, 2011. ISSN 1420-0597. doi:10.1007/s10596-010-9188-0.

[208] W. Kierat and U. Sztaba. *Distributions, Integral Transforms and Applications*. Analytical Methods and Special Functions. Taylor & Francis, Abingdon, UK, 2003.

[209] G. Kitagawa. A nonlinear smoothing method for time series analysis. *Statistica Sinica*, 1(2):371–388, July 1991.

[210] G. Kitagawa. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25, 1996. URL `http://www.jstor.org/stable/1390750`.

[211] D. C. Knill and A. Pouget. The Bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, 27(12):712–719, 2004. doi:10.1016/j.tins.2004.10.007.

[212] A. N. Kolmogorov. *Foundations of the Theory of Probability*. Chelsea Publishing Company, New York, second english edition, 1956.

[213] A. N. Kolmogorov. Interpolation und Extrapolation von stationären zufälligen Folgen. *Bull. Acad. Sci. URSS, Ser. Math.*, 5(1):3–14, 1941. URL `http://mi.mathnet.ru/eng/izv3775`.

[214] D. A. Kopriva. *Implementing Spectral Methods for Partial Differential Equations: Algorithms for Scientists and Engineers*. Scientific Computation. Springer-Verlag, Berlin, Germany, 2009. doi:10.1007/978-90-481-2261-5.

[215] J. Kotecha and P. Djuric. Gaussian sum particle filtering. *IEEE Transactions on Signal Processing*, 51(10):2602–2612, 2003. ISSN 1053-587X. doi:10.1109/TSP.2003.816754.

[216] A. Kovalenko, T. Mannseth, and G. Nævdal. Sampling error distribution for the ensemble Kalman filter update step. *Computational Geosciences*, Online:1–12, 2011. ISSN 1420-0597. doi:10.1007/s10596-011-9238-2.

[217] M. Krosche and R. Niekamp. Low rank approximation in spectral stochastic finite element method with solution space adaption. Technical Report 2010-3, Institut für Wissenschaftliches Rechnen, Technische Universität Braunschweig, Germany, November 2010. URL `http://www.digibib.tu-bs.de/?docid=00036351`.

[218] S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, March 1951. URL `http://www.jstor.org/stable/2236703`.

[219] A. Kučerová and H. G. Matthies. Uncertainty updating in the description of heterogeneous materials. *Technische Mechanik*, 30:211–226, 2010. URL `http://www.uni-magdeburg.de/ifme/zeitschrift_tm/2010_Heft1_3/18_Kucerova.html`.

[220] W. Lahoz, B. Khattatov, and R. Ménard, editors. *Data Assimilation: Making Sense of Observations*. Springer-Verlag, Berlin, Germany, 2010. doi:10.1007/978-3-540-74703-1.

[221] S. Lakshmivarahan and D. Stensrud. Ensemble Kalman filter. *IEEE Control Systems Magazine*, 29(3):34–46, 2009. ISSN 1066-033X. doi:10.1109/MCS.2009.932225.

[222] P. S. Laplace. Mémoir sur la probabilité des causes par les évènemens. *Laplace's 1774 Memoir in Inverse Probability, Statistical Science*, 1:364–378, 1774. doi:10.1214/ss/1177013621.

[223] K. J. H. Law and A. M. Stuart. Evaluating data assimilation algorithms. July 2011, `arXiv:1107.4118`.

[224] O. P. Le Maître and O. M. Knio. *Spectral Methods for Uncertainty Quantification with Applications to Computational Fluid Dynamics*. Scientific Computation. Springer, 2010. doi:10.1007/978-90-481-3520-2.

[225] R. Lebrun and A. Dutfoy. Do Rosenblatt and Nataf isoprob-

abilistic transformations really differ? *Probabilistic Engineering Mechanics*, 24(4):577–584, October 2009. ISSN 0266-8920. doi:10.1016/j.probengmech.2009.04.006.

[226] J. M. Lee. *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, Germany, 2002.

[227] E. L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer Texts in Statistics. Springer-Verlag, Berlin, Germany, second edition, 1998.

[228] F. Lehmann. Deterministic particle filtering for GPS navigation in the presence of multipath. *AEU – International Journal of Electronics and Communications*, 63(11):939–949, November 2009. ISSN 1434-8411. doi:10.1016/j.aeue.2008.07.009.

[229] C. Lemieux. *Monte Carlo and Quasi-Monte Carlo Sampling*. Springer Series in Statistics. Springer-Verlag, Berlin, Germany, 2009. doi:10.1007/978-0-387-78165-5.

[230] C. Lenglet, M. Rousson, R. Deriche, and O. Faugeras. Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor MRI processing. *Journal of Mathematical Imaging and Vision*, 25(3):423–444, 2006. ISSN 0924-9907. doi:10.1007/s10851-006-6897-z.

[231] P. F. J. Lermusiaux. Data assimilation via error subspace statistical estimation - part II: Middle Atlantic bight shelfbreak front simulations and ESSE validation. *Monthly Weather Review*, 127(7):1408–1432, July 1999. doi:10.1175/1520-0493(1999)127<1408:DAVESS>2.0.CO;2.

[232] P. F. J. Lermusiaux and A. R. Robinson. Data assimilation via error subspace statistical estimation - part I: Theory and schemes. *Monthly Weather Review*, 127(7):1385–1407, July 1999. doi:10.1175/1520-0493(1999)127<1385:DAVESS>2.0.CO;2.

[233] H. Li, J. Liu, and E. Kalnay. Correction of 'estimating observation impact without adjoint model in an ensemble Kalman filter'. *Quarterly Journal of the Royal Meteorological Society*, 136(651): 1652–1654, 2010. ISSN 1477-870X. doi:10.1002/qj.658.

[234] I. Lira. *Evaluating the Measurement Uncertainty—Fundamentals and Practical Guidance*. Series in Measurement and Technology. IOP Publishing Ltd., Bristol, Philadelphia, 2002.

[235] D. Liu. *Uncertainty Quantification with Shallow Water Equations*. PhD thesis, Carl-Friedrich-Gauss Faculty, University of Braunschweig and Faculty of Engineering, University of Florence, 2010.

[236] D. S. Liu and H. G. Matthies. Uncertainty quantification with spectral approximations of a flood model. *IOP Conference Series: Materials Science and Engineering*, 10(1):012208–, 2010. ISSN 1757-899X. doi:10.1088/1757-899X/10/1/012208.

[237] J. S. Liu, F. Liang, and W. H. Wong. The multiple-try method and local optimization in Metropolis sampling. *Journal of the American Statistical Association*, 95(449):121–134, 2000. URL `http://www.jstor.org/stable/2669532`.

[238] J. Liu and E. Kalnay. Estimating observation impact without adjoint model in an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 134(634):1327–1335, 2008. ISSN 1477-870X. doi:10.1002/qj.280.

[239] J. Liu, E. Kalnay, T. Miyoshi, and C. Cardinali. Analysis sensitivity calculation in an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 135(644):1842–1851, 2009. ISSN 1477-870X. doi:10.1002/qj.511.

[240] D. M. Livings, S. L. Dance, and N. K. Nichols. Unbiased ensemble square root filters. *Physica D: Nonlinear Phenomena*, 237(8):1021–1028, June 2008. ISSN 0167-2789. doi:10.1016/j.physd.2008.01.005.

[241] A. C. Lorenc. The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 129(595):3183–3203, 2003. ISSN 1477-870X. doi:10.1256/qj.02.132.

[242] E. N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, March 1963. doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

[243] S. Lototsky, R. Mikulevicius, and B. L. Rozovskii. Nonlinear filtering revisited: A spectral approach. *SIAM Journal on Control and Optimization*, 35(2):435–461, March 1997. doi:10.1137/S0363012993248918.

[244] S. Lototsky. Wiener chaos and nonlinear filtering. *Applied Mathematics and Optimization*, 54(3):265–291, November 2006. ISSN 0095-4616. doi:10.1007/s00245-006-0871-4.

[245] D. G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, Chichester, 1969.

[246] X. Luo, I. Moroz, and I. Hoteit. Scaled unscented transform Gaussian sum filter: Theory and application. *Physica D: Nonlinear Phenomena*, 239(10):684–701, May 2010. ISSN 0167-2789. doi:10.1016/j.physd.2010.01.022.

[247] P. Malliavin. *Stochastic Analysis*. Springer-Verlag, Berlin, Springer, 1997.

[248] J. Mandel, J. Beezley, J. Coen, and M. Kim. Data assimilation for wildland fires. *IEEE Control Systems Magazine*, 29(3):47–65, 2009. ISSN 1066-033X. doi:10.1109/MCS.2009.932224.

[249] J. Mandel. Efficient implementation of the ensemble Kalman filter. Technical Report UCDHSC/CCM Report No. 231, University of Colorado at Denver and Health Sciences Center, 2006.

[250] J. Mandel. A brief tutorial on the ensemble Kalman filter. Technical Report UCDHSC/CCM Report No. 242, University of Colorado at Denver and Health Sciences Center, 2007.

[251] J. Mandel, L. S. Bennethum, J. D. Beezley, J. L. Coen, C. C. Douglas, M. Kim, and A. Vodacek. A wildland fire model with data assimilation. *Mathematics and Computers in Simulation*, 79(3):584–606, December 2008. ISSN 0378-4754. doi:10.1016/j.matcom.2008.03.015.

[252] C.-M. Maris. Distance-based parameterisation and history matching. Master's thesis, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of

Technology, 2009. URL http://resolver.tudelft.nl/uuid:2cb210c7-4101-4fac-a82d-333fc0b8e0e8.

[253] Y. M. Marzouk and H. N. Najm. Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *Journal of Computational Physics*, 228(6):1862–1902, April 2009. ISSN 0021-9991. doi:10.1016/j.jcp.2008.11.024.

[254] Y. M. Marzouk and D. Xiu. A stochastic collocation approach to Bayesian inference in inverse problems. *Communications in Computational Physics*, 6(4):826–847, 2009.

[255] Y. M. Marzouk, H. N. Najm, and L. A. Rahn. Stochastic spectral methods for efficient Bayesian solution of inverse problems. *Journal of Computational Physics*, 224(2):560–586, June 2007. ISSN 0021-9991. doi:10.1016/j.jcp.2006.10.010.

[256] H. G. Matthies. *Encyclopedia of Computational Mechanics*, chapter Uncertainty Quantification with Stochastic Finite Elements. John Wiley & Sons, Chichester, 2007. doi:10.1002/0470091355.ecm071.

[257] H. G. Matthies. Stochastic finite elements: Computational approaches to stochastic partial differential equations. *ZAMM*, 88(11):849–873, 2008. doi:10.1002/zamm.200800095.

[258] H. G. Matthies. A hitchiker's guide to notation and definitions. Technical Report Informatikbericht 2011-05, Institut für Wissenschaftliches Rechnen, Technische Universität Braunschweig, August 2011. URL http://www.digibib.tu-bs.de/?docid=00040940.

[259] H. G. Matthies. White noise analysis for stochastic partial differential equations. Technical Report Informatikbericht 2011-03, Institut für Wissenschaftliches Rechnen, Technische Universität Braunschweig, 2011. URL http://www.digibib.tu-bs.de/?docid=00040820.

[260] D. McLaughlin and L. R. Townley. A reassessment of the groundwater inverse problem. *Water Resources Research*, 32(5):1131–1161, 1996. ISSN 0043-1397. doi:10.1029/96WR00160.

[261] N. Metropolis and S. Ulam. The Monte Carlo method. *Journal of the American Statistical Association*, 44(247):335–341, September 1949. URL `http://www.jstor.org/stable/2280232`.

[262] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953. doi:10.1063/1.1699114.

[263] R. N. Miller, M. Ghil, and F. Gauthiez. Advanced data assimilation in strongly nonlinear dynamical systems. *Journal of the Atmospheric Sciences*, 51(8):1037–1056, April 1994. doi:10.1175/1520-0469(1994)051<1037:ADAISN>2.0.CO;2.

[264] R. N. Miller, E. F. Carter, and S. T. Blue. Data assimilation into nonlinear stochastic models. *Tellus A*, 51(2):167–194, 1999. doi:10.1034/j.1600-0870.1999.t01-2-00002.x.

[265] T. Minka. Bayesian inference in dynamic models—an overview. Online, March 2007. URL `http://research.microsoft.com/en-us/um/people/minka/papers/dynamic.html`. Retrieved on 2012-01-15.

[266] H. L. Mitchell. Intercomparisons of 4D-variational assimilation and the ensemble Kalman filter. Online, 2010. URL `http://journals.ametsoc.org/page/Ensemble_Kalman_Filter`.

[267] H. L. Mitchell and P. L. Houtekamer. Ensemble Kalman filter configurations and their performance with the logistic map. *Monthly Weather Review*, 137(12):4325–4343, December 2009. doi:10.1175/2009MWR2823.1.

[268] H. L. Mitchell, P. L. Houtekamer, and G. Pellerin. Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Monthly Weather Review*, 130(11):2791–2808, November 2002. doi:10.1175/1520-0493(2002)130<2791:ESBAME>2.0.CO;2.

[269] T. Miyoshi. A review of the deterministic emsemble Kalman filtering methods and related techniques. Master's thesis, University of Maryland, Department of Meterology, 2004. URL `http://dods.atmos.umd.edu/theses_archive/2004/miyoshi/miyoshi.pdf`.

[270] T. Miyoshi. The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Monthly Weather Review*, 139(5):1519–1535, 2011. doi:10.1175/2010MWR3570.1.

[271] M. Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 26(3):735–747, January 2005. doi:10.1137/S0895479803436937.

[272] I. Myrseth and H. Omre. Hierarchical ensemble Kalman filter. *SPE Journal*, 15(2):569–580, June 2010. doi:10.2118/125851-PA.

[273] I. Myrseth, J. Sætrom, and H. Omre. Resampling the ensemble Kalman filter. Submitted for publication, 2010. URL `http://www.math.ntnu.no/~joeid/ure/resampling.pdf`.

[274] P. Naveau, M. G. Genton, and X. Shen. A skewed Kalman filter. *Journal of Multivariate Analysis*, 94(2):382–400, June 2005. ISSN 0047-259X. doi:10.1016/j.jmva.2004.06.002.

[275] L. Nerger, W. Hiller, and J. Schröter. A comparison of error subspace Kalman filters. *Tellus A*, 57:715–735, 2005. doi:10.1111/j.1600-0870.2005.00141.x.

[276] L. Nerger, S. Danilov, W. Hiller, and J. Schröter. Using sea-level data to constrain a finite-element primitive-equation ocean model with a local SEIK filter. *Ocean Dynamics*, 56(5):634–649, 2006. ISSN 1616-7341. doi:10.1007/s10236-006-0083-0.

[277] G.-H. C. Ng, D. McLaughlin, D. Entekhabi, and A. Ahanin. The role of model dynamics in ensemble Kalman filter performance for chaotic systems. *Tellus A*, 63(5):958–977, 2011. ISSN 1600-0870. doi:10.1111/j.1600-0870.2011.00539.x.

[278] A. Nouy and O. P. Le Maître. Generalized spectral decomposition for stochastic nonlinear problems. *Journal of Computational Physics*, 228(1):202–235, January 2009. ISSN 0021-9991. doi:10.1016/j.jcp.2008.09.010.

[279] G. Nævdal, R. Hanea, D. Oliver, and B. Vallès. Ensemble

Kalman filter for model updating – a special issue. *Computational Geosciences*, 15(2):223–224, 2011. ISSN 1420-0597. doi:10.1007/s10596-011-9228-4.

[280] J. T. Oden. Historical comments on finite elements. In *A History of Scientific Computing*, pages 152–166. Addison-Wesley, 1990.

[281] P. Oke, P. Sakov, and S. Corney. Impacts of localisation in the EnKF and EnOI: experiments with a small model. *Ocean Dynamics*, 57 (1):32–45, February 2007. doi:10.1007/s10236-006-0088-8.

[282] B. K. Øksendal. *Stochastic Differential Equations—An Introduction with Applications*. Springer-Verlag, Berlin, Germany, sixth edition, 2003.

[283] D. Oliver and Y. Chen. Improved initial sampling for the ensemble Kalman filter. *Computational Geosciences*, 13(1):13–27, 2009. ISSN 1420-0597. doi:10.1007/s10596-008-9101-2.

[284] D. Oliver and Y. Chen. Recent progress on reservoir history matching: a review. *Computational Geosciences*, 15(1):185–221, 2011. ISSN 1420-0597. doi:10.1007/s10596-010-9194-2.

[285] E. Ott, B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A*, 56(5):415–428, 2004. doi:10.1111/j.1600-0870.2004.00076.x.

[286] O. Pannekoucke, L. Berre, and G. Desroziers. Filtering properties of wavelets for local background-error correlations. *Quarterly Journal of the Royal Meteorological Society*, 133(623):363–379, 2007. doi:10.1002/qj.33.

[287] M. L. Parks, E. de Sturler, G. Mackey, D. D. Johnson, and S. Maiti. Recycling Krylov subspaces for sequences of linear systems. *SIAM Journal on Scientific Computing*, 28(5):1651–1674, January 2006. doi:10.1137/040607277.

[288] K. B. Petersen and M. S. Pedersen. The matrix cookbook. PDF, November 2008. URL `http://matrixcookbook.com`.

[289] R. E. Petrie. Localization in the ensemble Kalman filter. Master's thesis, Department of Meteorology, University of Reading, August 2008. URL `http://www.met.reading.ac.uk/~wx019276/documents/Thesis.pdf`.

[290] H. A. Phale and D. S. Oliver. Data assimilation using the constrained ensemble Kalman filter. *SPE Journal*, 16(2):331–342, 2011. doi:10.2118/125101-PA.

[291] D. T. Pham. Dimension, predictability and reduced rank Kalman filtering in data assimilation. In *Proceedings of the Third Bilateral French-Russian Conference: Predictability of Atmospheric and Oceanic Circulations*. Institute Franco-Russe A. M. Lyapunov (INRA - Moscow State University), 1997.

[292] D. T. Pham. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly Weather Review*, 129(5):1194–1207, May 2001. doi:10.1175/1520-0493(2001)129<1194:SMFSDA>2.0.CO;2.

[293] D. T. Pham, J. Verron, and M. Christine Roubaud. A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine Systems*, 16(3-4):323–340, October 1998. ISSN 0924-7963. doi:10.1016/S0924-7963(97)00109-7.

[294] D. J. Posselt and C. H. Bishop. Nonlinear parameter estimation: Comparison of an ensemble Kalman smoother with a Markov chain Monte Carlo algorithm. *Monthly Weather Review*, 140(6):1957–1974, 2012. ISSN 0027-0644. doi:10.1175/MWR-D-11-00242.1.

[295] J. E. Potter and R. G. Stern. Statistical filtering of space navigation measurements. In *Proceedings of the AIAA Guidance and Control Conference, Massachusetts Institute of Technology, August*, 1963.

[296] W. B. Powell. *Approximate Dynamic Programming*. John Wiley & Sons, 2007.

[297] F. Rabier, H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons. The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified phys-

ics. *Quarterly Journal of the Royal Meteorological Society*, 126(564): 1143–1170, 2000. doi:10.1002/qj.49712656415.

[298] I. Rechenberg. *Evolution Strategy: Optimizations of Technical Systems Based on Principles of Biological Evolution*. Frommann-Holzboog, Stuttgart, 1972.

[299] R. H. Reichle, D. B. McLaughlin, and D. Entekhabi. Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, 130(1):103–114, 2002. doi:10.1175/1520-0493(2002)130<0103%3AHDAWTE>2.0.CO%3B2.

[300] N. Remy, A. Boucher, and J. Wu. *Applied Geostatistics with SGeMS*. Cambridge University Press, Cambridge, UK, 2009.

[301] A. Reynolds, N. He, L. Chu, and D. Oliver. Reparameterization techniques for generating reservoir descriptions conditioned to variograms and well-test pressure data. *SPE Journal*, 1(4):413–426, 1996. doi:10.2118/30588-PA.

[302] H. Risken. *The Fokker-Planck Equation: Methods of Solution and Applications*. Springer Series in Synergetics. Springer-Verlag, Berlin, Germany, second edition, 1989.

[303] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman filter : particle filters for tracking applications*. Artech House Inc., Boston, 2004.

[304] R. Rojas. *Neural Networks: A Systematic Introduction*. Springer-Verlag, Berlin, Germany, 1996.

[305] J. S. Rosenthal. *A first look at rigorous probability theory*. World Scientific Publishing Co. Pte. Ltd., Singapore, second edition, 2006.

[306] D. Rozier, F. Birol, E. Cosme, P. Brasseur, J. M. Brankart, and J. Verron. A reduced-order Kalman filter for data assimilation in physical oceanography. *SIAM Review*, 49(3):449–465, January 2007. doi:10.1137/050635717.

[307] G. A. Saad. *Stochastic Data Assimilation with Application to Multi-*

Phase Flow and Health Monitoring Problems. PhD thesis, Faculty of the Graduate School, University of Southern California, 2007.

[308] P. Sakov and L. Bertino. Relation between two common localisation methods for the EnKF. *Computational Geosciences*, 15(2):225–237, 2011. ISSN 1420-0597. doi:10.1007/s10596-010-9202-6.

[309] P. Sakov and L. Bertino. Non-equivalence of parallel and serial assimilation with EnKF in presence of localisation. Presentation at eVITA meeting, Geilo, Norway, January 11-16, 2009. URL `http://msc.nersc.no/files/sakov_bertino_evita2009.pdf`.

[310] P. Sakov and P. R. Oke. Implications of the form of the ensemble transformation in the ensemble square root filters. *Monthly Weather Review*, 136(3):1042–1053, March 2008. doi:10.1175/2007MWR2021.1.

[311] P. Sakov and P. R. Oke. A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. *Tellus A*, 60(2):361–371, 2008. doi:10.1111/j.1600-0870.2007.00299.x.

[312] P. Sakov, D. Oliver, and L. Bertino. An iterative EnKF for strongly nonlinear systems. *Monthly Weather Review*, 140(6):1988–2004, 2012. ISSN 0027-0644. doi:10.1175/MWR-D-11-00176.1.

[313] R. Schulze-Riegert and S. Ghedan. Modern techniques for history matching. In *9th International Forum on Reservoir Simulation, December 9 – 13, Abu Dhabi, UAE*, 2007.

[314] H.-P. Schwefel. Kybernetische Evolution als Strategie der experimentellen Forschung in der Strömungstechnik. Master's thesis, Technische Universitat Berlin, March 1965.

[315] H.-P. Schwefel. *Evolutionsstrategie und numerische Optimierung*. PhD thesis, Technische Universität Berlin, 1975.

[316] D. W. Scott and S. R. Sain. Multidimensional density estimation. In E. W. C.R. Rao and J. Solka, editors, *Data Mining and Data Visualization*, volume Volume 24, pages 229–261. Elsevier, 2005. doi:10.1016/S0169-7161(04)24009-3.

[317] I. E. Segal and R. A. Kunze. *Integrals and Operators*. Springer-Verlag, Berlin, Germany, 1978.

[318] D. Seleši. Hilbert space valued generalized random processes—part I. *Novi Sad Journal of Mathematics*, 37(1):129–154, 2007.

[319] D. Seleši. *Generalized Stochastic Processes in Infinite Dimensional Spaces with Applications to Singular Stochastic Partial Differential Equations*. PhD thesis, University of Novi Sad, Faculty of Science, Department of Mathematics and Informatics, 2007.

[320] M. Sen and P. L. Stoffa. *Global Optimization Methods in Geophysical Inversion*. Elsevier, Amsterdam, The Netherlands, 1995.

[321] C. Y. Shen, T. E. Evans, and S. Finette. Polynomial chaos quantification of the growth of uncertainty investigated with a Lorenz model. *Journal of Atmospheric and Oceanic Technology*, 27(6):1059–1071, 2010. doi:10.1175/2009JTECHO727.1.

[322] N. Shinozaki, M. Sibuya, and K. Tanabe. Numerical algorithms for the Moore-Penrose inverse of a matrix: direct methods. *Annals of the Institute of Statistical Mathematics*, 24(1):193–203, 1972. ISSN 0020-3157. doi:10.1007/BF02479751.

[323] A. V. Skorokhod. *Studies in the Theory of Random Processes*. Dover Publications, New York, 1982.

[324] K. W. Smith. Cluster ensemble Kalman filter. *Tellus A*, 59(5):749–757, 2007. ISSN 1600-0870. doi:10.1111/j.1600-0870.2007.00246.x.

[325] C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson. Obstacles to high-dimensional particle filtering. *Monthly Weather Review*, 136 (12):4629–4640, December 2008. doi:10.1175/2008MWR2529.1.

[326] D. Sornette and K. Ide. The Kalman-Lévy filter. *Physica D: Nonlinear Phenomena*, 151(2-4):142–174, May 2001. ISSN 0167-2789. doi:10.1016/S0167-2789(01)00228-7.

[327] J. C. Spall. *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. John Wiley & Sons, Chichester, April 2003. ISBN 978-0-471-33052-3.

[328] M. D. Springer. *The algebra of random variables*. Probability and Statistics Series. Wiley, 1979. ISBN 9780471014065.

[329] R. Steinacker, D. Mayer, and A. Steiner. Data quality control based on self-consistency. *Monthly Weather Review*, 139(12):3974–3991, August 2011. ISSN 0027-0644. doi:10.1175/MWR-D-10-05024.1.

[330] A. Stordal, H. Karlsen, G. Nævdal, H. Skaug, and B. Vallès. Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter. *Computational Geosciences*, 15(2): 293–305, 2011. ISSN 1420-0597. doi:10.1007/s10596-010-9207-1.

[331] R. L. Stratonovich. Conditional Markov processes. *Theory of Probability and its Applications*, 5(2):156–178, 1960. doi:10.1137/1105015.

[332] S. Strebelle. Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology*, 34(1):1–21, 2002. ISSN 0882-8121. doi:10.1023/A:1014009426274.

[333] S. B. Strebelle and A. G. Journel. Reservoir modeling using multiple-point statistics. In *SPE Annual Technical Conference and Exhibition, 30 September-3 October, New Orleans, Louisiana*, 2001. doi:10.2118/71324-MS.

[334] D. W. Stroock. *An Introduction to Markov Processes*, volume 230 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, Germany, 2005.

[335] A. M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559, 2010. doi:10.1017/S0962492910000061.

[336] J. Sætrom and H. Omre. Ensemble Kalman filtering for nonlinear likelihood models using kernel-shrinkage regression techniques. *Computational Geosciences*, 15(3):529–544, 2011. ISSN 1420-0597. doi:10.1007/s10596-010-9222-2.

[337] J. Sætrom and H. Omre. Ensemble Kalman filtering with shrinkage regression techniques. *Computational Geosciences*, 15(2):271–292, 2011. ISSN 1420-0597. doi:10.1007/s10596-010-9196-0.

[338] B. G. Tabachnick and L. S. Fidell. *Using Multivariate Statistics*. Pearson, fifth edition, 2007.

[339] Y. Tang and J. Ambadan. Reply. *Journal of the Atmospheric Sciences*, 66(11):3501–3503, November 2009. ISSN 0022-4928. doi:10.1175/2009JAS3273.1.

[340] A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, Philadelphia, USA, 2005.

[341] M. K. Tippett, J. L. Anderson, and C. H. Bishop. Ensemble square root filters. *Monthly Weather Review*, 131:1485 – 1490, 2003. doi:10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2.

[342] S. Torquato. *Random Heterogeneous Materials: Microstructure and Macroscopic Properties*. Springer-Verlag, Berlin, 2002.

[343] G. Ueno and T. Tsuchiya. Covariance regularization in inverse space. *Quarterly Journal of the Royal Meteorological Society*, 135 (642):1133–1156, 2009. ISSN 1477-870X. doi:10.1002/qj.445.

[344] F. van der Heijden, R. Duin, D. de Ridder, and D. M. J. Tax. *Classification, Parameter Estimation and State Estimation: An Engineering Approach using MATLAB*. John Wiley & Sons, Chichester, 2004.

[345] R. van der Merwe. *Sigma-Point Kalman Filters for Probabilistic Inference in Dynamic State-Space Models*. PhD thesis, OGI School of Science & Engineering at Oregon Health & Science University, 2004.

[346] R. van der Merwe and E. Wan. Gaussian mixture sigma-point particle filters for sequential probabilistic inference in dynamic state-space models. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 6, pages VI 701–704, 2003. doi:10.1109/ICASSP.2003.1201778.

[347] R. van der Merwe, A. Doucet, N. de Freitas, and E. Wan. The unscented particle filter. Technical report, Cambridge University Engineering Department, 2000.

[348] P. J. van Leeuwen. Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Quarterly Journal of the Royal Meteorological Society*, 136(653):1991–1999, 2010. ISSN 1477-870X. doi:10.1002/qj.699.

[349] P. J. van Leeuwen. A variance-minimizing filter for large-scale applications. *Monthly Weather Review*, 131(9):2071–2084, September 2003. doi:10.1175/1520-0493(2003)131<2071:AVFFLA>2.0.CO;2.

[350] P. J. van Leeuwen. Particle filtering in geophysical systems. *Monthly Weather Review*, 137(12):4089–4114, December 2009. doi:10.1175/2009MWR2835.1.

[351] V. Černý. Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm. *Journal of Optimization Theory and Applications*, 45(1):41–51, January 1985. doi:10.1007/BF00940812.

[352] J. F. Vincent, O. A. Bogatyreva, N. R. Bogatyrev, A. Bowyer, and A.-K. Pahl. Biomimetics: its practice and theory. *Journal of The Royal Society Interface*, 3(9):471–482, 2006. doi:10.1098/rsif.2006.0127.

[353] E. Wan and R. Van Der Merwe. The unscented Kalman filter for nonlinear estimation. In *The IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium (AS-SPCC) 2000*, pages 153–158, 2000. doi:10.1109/ASSPCC.2000.882463.

[354] X. Wang, C. H. Bishop, and S. J. Julier. Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble? *Monthly Weather Review*, 132(7), 2004.

[355] E. W. Weisstein. Sample variance computation. From MathWorld–A Wolfram Web Resource, March 2011. URL `http://mathworld.wolfram.com/SampleVarianceComputation.html`.

[356] G. Welch and G. Bishop. An introduction to the Kalman filter. Technical report, Department of Computer Science, University of North Carolina at Chapel Hill, 2006. URL `http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf`.

[357] M. West. ISO 8000 — the emerging standard for data quality. *IDQ Newsletter*, 5(3):Online, 2009. URL `http://iaidq.org/publications/doc/west-2009-07.shtml`.

[358] J. S. Whitaker and T. M. Hamill. Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130(7):1913–1924, July 2002. doi:10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

[359] P. Whittle. *Probability via expectation*. Springer-Verlag, Berlin, Germany, fourth edition, 2000.

[360] N. Wiener. The homogeneous chaos. *American Journal of Mathematics*, 60(4):897–936, October 1938. URL `http://www.jstor.org/stable/2371268`.

[361] C. K. Wikle and L. M. Berliner. A Bayesian tutorial for data assimilation. *Physica D: Nonlinear Phenomena*, 230(1-2):1–16, June 2007. ISSN 0167-2789. doi:10.1016/j.physd.2006.09.017.

[362] D. Xiu. Fast numerical methods for stochastic computations: A review. *Communications in Computational Physics*, 5:242–272, 2009.

[363] D. Xiu. *Numerical Methods for Stochastic Computations—A Spectral Method Approach*. Princeton University Press, Princeton, NJ, 2010. URL `http://press.princeton.edu/titles/9229.html`.

[364] D. Xiu and G. E. Karniadakis. The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, 2002. doi:10.1137/S1064827501387826.

[365] S.-C. Yang and E. Kalnay. Handling nonlinearity and non-Gaussianity in the ensemble Kalman filter: Experiments with the three-variable Lorenz model. *Monthly Weather Review*, Submitted for publication, 2011.

[366] S.-C. Yang, E. Kalnay, B. Hunt, and N. E. Bowler. Weight interpolation for efficient data assimilation with the local ensemble transform Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 135(638):251–262, 2009. ISSN 1477-870X. doi:10.1002/qj.353.

[367] M. Zakai. On the optimal filtering of diffusion processes. *Probability Theory and Related Fields*, 11(3):230–243, 1969. ISSN 0178-8051. doi:10.1007/BF00536382.

[368] E. Zander and H. G. Matthies. Tensor product methods for stochastic problems. *PAMM*, 7:2040067 – 2040068, 2008. doi:10.1002/pamm.200700773.

[369] L. Zeng and D. Zhang. A stochastic collocation based Kalman filter for data assimilation. *Computational Geosciences*, 14(4):721–744, 2010. ISSN 1420-0597. doi:10.1007/s10596-010-9183-5.

[370] Y. Zhang and D. Oliver. Improving the ensemble estimate of the Kalman gain by bootstrap sampling. *Mathematical Geosciences*, 42 (3):327–345, 2010. ISSN 1874-8961. doi:10.1007/s11004-010-9267-8.

[371] Y. Zhang and D. Oliver. Evaluation and error analysis: Kalman gain regularization versus covariance regularization. *Computational Geosciences*, 15(3):489–508, 2011. ISSN 1420-0597. doi:10.1007/s10596-010-9218-y.

[372] Y. Zhang, N. Liu, and D. Oliver. Ensemble filter methods with perturbed observations applied to nonlinear problems. *Computational Geosciences*, 14(2):249–261, 2010. ISSN 1420-0597. doi:10.1007/s10596-009-9149-7.

[373] H. Zhou, J. J. Gómez-Hernández, H.-J. Hendricks Franssen, and L. Li. An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering. *Advances in Water Resources*, 34(7):844–864, July 2011. ISSN 0309-1708. doi:10.1016/j.advwatres.2011.04.014.

[374] M. Zupanski. Maximum likelihood ensemble filter: Theoretical aspects. *Monthly Weather Review*, 133(6):1710–1726, June 2005. doi:10.1175/MWR2946.1.

[375] M. Zupanski, I. M. Navon, and D. Zupanski. The maximum likelihood ensemble filter as a non-differentiable minimization algorithm. *Quarterly Journal of the Royal Meteorological Society*, 134(633): 1039–1050, 2008. doi:10.1002/qj.251.